R E P O R T   R E S U M E S

ED 013 542   EM 000 524

EFFECTIVENESS OF TRAFFIC SAFETY FILMS IN RELATION TO EMOTIONAL INVOLVEMENT.

BY- LEVONIAN, EDWARD   CASE, HARRY W.
CALIFORNIA UNIV., LOS ANGELES

REPORT NUMBER ITTE-RR-40   PUB DATE   JUN 65
EDRS PRICE MF-$0.75 HC-$7.72   193P.

DESCRIPTORS- *HIGH SCHOOL STUDENTS, *DRIVER EDUCATION, *RETENTION, *ANXIETY, *INSTRUCTIONAL FILMS, FEAR, EFFECTIVE TEACHING, TRAFFIC SAFETY, LOS ANGELES, CALIFORNIA

HIGH SCHOOL STUDENTS ENROLLED IN FOUR DRIVER EDUCATION CLASSES WERE SHOWN A TRAFFIC SAFETY FILM, THEN TESTED FOR INFORMATION RETENTION AFTER 10 MINUTES AND AGAIN AFTER 1 WEEK. FORGETTING WAS DEFINED AS A CORRECT RESPONSE ON THE FIRST TEST BUT NOT ON THE SECOND, WHILE REMINISCENCE WAS DEFINED AS THE CONVERSE. RETENTION WAS DEFINED AS A CORRECT RESPONSE BOTH TIMES. ANXIETY AROUSAL DURING FILM PRESENTATION WAS MEASURED VIA GALVANIC SKIN RESPONSE AND PLOTTED THROUGH SUCCESSIVE MINUTES OF THE FILM. THE PRIMARY HYPOTHESIS THAT FORGETTING, RETENTION, AND REMINISCENCE WOULD BE ASSOCIATED WITH SMALLER, MEDIUM, AND LARGER INCREASES IN AROUSAL NEAR THE MOMENT OF PRESENTATION WAS SUPPORTED. THIS FINDING RUNS COUNTER TO A BELIEF OFTEN EXPRESSED IN THE LITERATURE (BUT, AS POINTED OUT IN A LENGTHY REVIEW IN THIS REPORT, NOT SUPPORTED BY EVIDENCE) THAT COMMUNICATION EFFECTIVENESS DEPENDS ON A LOW LEVEL OF ANXIETY AROUSAL OR AN AROUSAL REDUCTION. THESE RESULTS IMPLY THAT TRAFFIC SAFETY FILMS (FOR INSTANCE) MAY BE MORE EFFECTIVE IF ANXIETY OR FEAR IS AROUSED. A COROLLARY CONCLUSION IS THAT AN EFFECTIVE FILM SHOULD ALTERNATE SEQUENCES THAT AROUSE AND REDUCE ANXIETY, WITH INFORMATION PRESENTED AT MOMENTS OF ANXIETY AROUSAL ONLY. CONTENT OF ANXIETY-AROUSING SEQUENCES NEED NOT BE PERTINENT TO INFORMATION TO BE RETAINED. ON A METHODOLOGICAL ISSUE, THE AUTHORS STRESS THE NECESSITY FOR A DELAYED TEST OF INFORMATION RETENTION, BECAUSE OF THE REMINISCENCE PHENOMENON. (GO)

ED013542

Research Report 40

June 1965

A

EFFECTIVENESS OF TRAFFIC SAFETY FILMS IN RELATION TO EMOTIONAL INVOLVEMENT

By

Edward Levonian

EM 000524

THE INSTITUTE OF TRANSPORTATION
AND TRAFFIC ENGINEERING
UNIVERSITY OF CALIFORNIA
LOS ANGELES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EFFECTIVENESS OF TRAFFIC SAFETY FILMS IN RELATION TO EMOTIONAL INVOLVEMENT

Edward Levonian

Institute of Transportation and Traffic Engineering
University of California, Los Angeles

June, 1965

EM 000 534

Acknowledgements

This research was supported by the Highway Transportation Agency, State of California, and the Bureau of Public Roads, U. S. Department of Commerce, under a contract with the University of California. Harry W. Case, Associate Director, ITTE, was Principal Investigator.

As in many previous studies, a number of people in the Los Angeles City School System have given complete cooperation in arranging for the data collection. Particular appreciation is expressed to Dr. Gerwin C. Neher, Secondary Administrative Coordinator, Dr. Herbert Popenoe, Chairman, Committee on Research Studies, and Mr. Melvin T. Schroeder, Supervisor, Driver Instruction, for their facilitation of arrangements, to Mr. Robert Lewis, Safety Specialist, and Dr. Eugene Olson, Principal, University High School, for their suggestions in local arrangements, and to Mr. George L. Wilson and Mr. Jack Moscowitz for their cooperation in the collection of data in their driver education classes.

Considerable contributions were made by Mr. Stephen Canfield, Mr. Michael Kaufman, and Miss Carolee Wiles throughout the data collection and processing phases of this study.

All computations were performed at the UCLA Computing Facility.

Contents

1 Problem 1
  1.1 Research on Communication Effectiveness and Arousal .. 4
    1.11 Studies Cited in Hovland, Janis, and Kelley 5
    1.12 Janis and Feshbach (1953) 8
    1.13 Moltz and Thistlethwaite (1955) 46
    1.14 Goldstein (1959) 49
    1.15 Berkowitz and Cottingham 63
    1.16 Janis and Terwilliger (1962) 69
    1.17 Snider (1962) 74
    1.18 Summary 77
  1.2 Research on Learning and Remembering 85
    1.21 Learning and the Orienting Response 86
    1.22 Remembering and Arousal 89
  1.3 Implications for Current Study 95
    1.31 Theoretical Implications 95
    1.32 Methodological Implications 106
2 Method 111
  2.1 Subjects 111
  2.2 Procedure 111
  2.3 Questionnaire 112
  2.4 Film 114
  2.5 Skin Resistance 115
3 Results 117
  3.1 Type of Event in Relation to Decrement 118
  3.2 Temporal Relation of Decrement to Presentation 125
4 Discussion 129
  4.1 Information Accessibility 129
  4.2 Biological Considerations 134
  4.3 Comparison with Section 1 Studies 142
  4.4 General Implications of the Results 165
  4.5 Implications for Traffic Safety Films 169
5 Summary 175
  References 177

1 Problem

The number of traffic safety film titles currently in use in the United States is about 500 (Merrill, 1959; American Insurance Association, 1965; National Safety Council, 1965; National Education Association, 1960). With each title involving as many as 2,000 prints (Malfetti, 1961), and with each print being seen on the average by perhaps 1,000 persons in classroom-size groups each year (Merrill, 1959), a reasonable estimate of the annual number of viewings of traffic safety films is perhaps 25,000,000.

Traffic safety communications utilize mass media other than the film medium. Popular pamphlets are distributed by the millions (Malfetti, 1961). Billboards, radio spots, and articles in popular magazines are common vehicles for the transmission of traffic safety messages. Such messages recorded on film and transmitted by television reach millions more (Creative Research Associates, 1961; Naisbitt, 1961); the public interest telecast which drew the largest viewing audience as of 1965 dealt with traffic safety, encouraging the continued distribution of the program in booklet form (National Drivers Test, 1965). Other material is available in programed text form (Ferraro, Teal, and Fabrizio, 1964; Harris, 1965; Rohloff, 1964; Velte, 1965).

Studies which have implications for the effectiveness of traffic safety films sometimes involve such films (Brody, 1961; McAshan, 1960; Merrill, 1962; Merrill and McAshan, 1960), but more often involve mass communications other than traffic safety films (Blumenthal, 1964). This is not inappropriate, for unless the results in a particular study are content-specific, then studies not involving traffic safety films may well have implications for traffic safety films; conversely, studies utilizing traffic safety films may have implications for communication effectiveness in general, independent of specific content. In general, broader implications are likely to accrue from studies involving variables which are not media-specific or content-specific.

It is assumed that a partial measure of the effectiveness of traffic safety films is the retention of information presented in the film. Since retention is the only measure of effectiveness used in this study, and since a complete evaluation of effectiveness should also take into account other measures, and in particular driving behavior as a validating measure, it should be recognized that this study makes only a partial assessment of traffic safety films.

If it is the intent of the communicator to bring about a nonverbal behavioral change in the communicatee, then the effectiveness of the communication should be assessed in terms of the intended behavioral change. A nonverbal behavioral measure of effectiveness was not used in the current study because it was deemed more important at this time to resolve a basic problem concerning the important parameters relating retention to emotional involvement. Once this problem is resolved, the relation between communication effectiveness, measured behaviorally, and emotional involvement can be investigated with a firmer background.

Research in human communication has generally assumed that retention of communication content is an important factor in communication effectiveness. Thus, many studies of communication effectiveness obtain a measure of information retention.

The term "emotional involvement" in this report pertains to the arousal (a term which will be used frequently for the sake of parsimony) of the subject while he is exposed to the film. Arousal is frequently measured physiologically, and quite a number of studies have measured physiological functions during picture viewing (Davis, 1957; Graham and Graham, 1964), or film viewing (Aas, 1958; Alexander, Flagg, Foster, Clemens, and Blahd, 1961; Becker, 1960; Becker, 1964;

Berkowitz, 1964; Berkowitz and Rawlings, 1963; Davison, 1963; Dysinger and Ruckmick, 1933; Gestant and Bert, 1954; Guba, Wolf, de Groot, Knemeyer, Van Atta, and Light, 1964; Humphrey, 1950; Kleitman, 1945; Lazarus, 1964; Lazarus, Speisman, Mordkoff, and Davison, 1962; Levonian, 1962; Maccoby, Wilson, and Burton, 1958; Mordkoff, 1963; Schwartz, 1956; Wendt, 1948). However, only one of these studies (Becker, 1964) involved the direct measurement of retention in relation to arousal during the film, and this study is discussed later.

The term "communication effectiveness" as used in this report pertains to the effectiveness of the communication in mediating those changes in the communicatee (e. g., student, viewer) which are desired by the communicator (e. g., author, producer). The term will be restricted to those cases in which the communicator intends or advocates a specified change for the communicatee. This restriction eliminates from consideration those studies which involve a measure of "communication effect" but not of communication effectiveness - that is, those studies in which the communication had an effect but one which was not that advocated by the communicator. For instance, a number of studies have demonstrated that films which show aggression have an effect on the aggressive behavior of subjects after they viewed the films (Bandura, Ross, and Ross, 1963; Berkowitz, 1964; Berkowitz and Rawlings, 1963; Lövass, 1961; Mussen and Rutherford, 1961; Walters, Thomas, and Acker, 1962). While this type of study indicates that emotion-inducing films influence subsequent behavior, the behavioral change cannot be taken as a measure of communication effectiveness because, presumably, this particular behavior was not specifically advocated by the producer. In summary, any effect of a communication can constitute a measure of communication effect, but only the specific effect intended by the communicator constitutes a measure of "communication effectiveness," as this term is used in this study. In the current study, arousal is a measure of communication effect, retention a measure of communication effectiveness.

Communication effectiveness in relation to arousal has been explored in a number of studies. Very often the investigator concludes that anxiety induced by the communication makes it less effective. On the other hand, a high proportion of traffic safety films intentionally induce anxiety by emphasizing the hazards of negligent driving under the assumption that such behavior will be eschewed (Malfetti, 1961).

The relative merit of these two positions may be elucidated to some extent by a review of the relevant research. The arguments against emotion-inducing communications tend to arise from mass communication research in social psychology. Arguments in favor of emotion-inducing communications tend to arise from learning-remembering research in experimental psychology. These two areas will be reviewed in the next two major sections, after which a presentation will be given of the extent to which the reviewed material has implications for the current study.

In order to facilitate comparisons among studies, each probability level presented in this report will be two tailed. The .05 significance level will be assumed for every study in this report except for the current study, for which the .01 level will be employed.

1.1 Research on Communication Effectiveness and Arousal

The studies in this section are presented chronologically. Such an organization not only helps to reveal temporal trends in methodology, results, and interpretations, but also articulates neatly with Section 1.2, which lists studies on perserverative consolidation and orienting response, studies which tend to be more recent.

Mass communication studies involving arousal have been directed primarily along conceptual lines since the pioneering study of Janis and Feshbach (1953). Many of the studies in the preceding period were directed to questions revolving around the relative effectiveness of emotional and logical mass communications. Three of these earlier studies are listed by Hovland, Janis, and Kelley (1953) in their book on communication and persuasion. A review of these three studies will suffice to indicate the earlier research in the area of communication effectiveness in relation to emotional involvement.

1.11 Studies Cited in Hovland, Janis, and Kelley

Hovland, Janis, and Kelley (1953), in their chapter on fear-arousing appeals, cite three early studies related to the effectiveness of emotional and logical appeals. They cite two studies (Hartmann, 1936; Menefee and Granneberg, 1940) as examples which reported emotional appeals as being more effective than rational appeals, and one study (Knower, 1935) which reported rational appeals as being more effective than emotional ones under certain conditions.

In the Hartmann study (1936), an emotional leaflet was distributed in one ward preceding an election in 1935, a rational leaflet in another ward. Both advocated voting for the Socialist candidate. The effectiveness of each leaflet was measured by the increase in votes for the Socialist candidate over that of the previous year. Hovland, Janis, and Kelley cite the fact that the Socialist vote increased by 50% in the "emotional" ward, but only 35% in the "rational" ward, as evidence of the relatively greater effectiveness of the emotional leaflet. They fail to indicate, however, that this difference in percentages was due entirely to 10 voters - five more Socialist votes in the "emotional" ward than would be expected on the basis of the 1934 Socialist vote in that ward, and five less Socialist votes in the "rational" ward than expected - and that the difference between obtained and expected frequencies is significant at only the .50 level. Hartmann also chose to ignore this fact, and concluded that "There seems to be no escape from the decision that the emotional political appeal is a better vote-getting instrument than the rational approach...." (Hartmann, 1936, p. 113).

In the Menefee and Granneberg study (1940), printed materials were read to two groups of university students. Both presentations advocated isolationism as the appropriate foreign policy for the United States, but one utilized an emotional appeal, the other a rational appeal. Immediately after the first reading, the material was reread slowly, and immediately after that, the students responded to 10 opinion items pertaining to isolationism. Students who heard the emotional appeal gave more isolationism responses, but the authors failed to make a significance test of the difference between these two groups with respect to these 10 items, and insufficient data are presented in the article to allow the reader to determine the exact significance level of the difference. However, the data available indicate that the difference is not likely to reach the .20 level of significance. Menefee and Granneberg also presented the alternative argument (collective security) to two additional groups, and again the emotional appeal was more effective, but again it is impossible to determine the exact level of significance of the difference. However, the data available indicate that the difference is not likely to reach the .40 level. After responding to the 10 opinion items, the students responded to a criterion item: "Which type of foreign policy do you prefer for the United States: isolation or collective security?" The difference in response to this criterion item can be tested exactly for significance. The difference between the groups which heard the isolationism argument was in favor of the emotional appeal, but significant at only the .98 level. The difference between the groups which heard the collective security was in favor of the rational appeal, but significant at only the .52 level. Despite the fact that their data fail to give clear support for either type of appeal, Menefee and Granneberg conclude that in their study ". . . . emotional propaganda was much more effective in changing student opinion regarding our foreign policy than was logical argumentation."

In the Knower study (1935), two speeches were presented to two groups of university students. Both presentations advocated prohibition, but one utilized an emotional appeal, the other a logical appeal. An attitude toward prohibition questionnaire was administered 2 to 6 weeks before the presentations, and the same attitude questionnaire was administered immediately after the speech. On the basis of the initial questionnaire, wet students were identified and exposed to emotional or logical versions of the prohibition speech. Each version resulted in an attitude change in the intended direction, but the difference between the two changes, which was in favor of the logical appeal, was significant at only the .20 level. Knower also presented the alternative argument (anti-prohibition) to the dry students, with half of them receiving the emotional appeal, the other half of them the logical appeal. Again, both groups changed in the intended direction, but the difference between the two changes, this time in favor of the emotional appeal, was significant at only the .95 level.

The last three paragraphs have reviewed the three studies cited by Hovland, Janis, and Kelley (1953) as bearing on the problem of the effectiveness of emotional and logical communications. Two of them (Hartmann, 1936; Menefee and Granneberg, 1940) were cited as yielding evidence on the greater effectiveness of emotional communications, while the third (Knower, 1935) was cited as favoring logical communications, at least under certain conditions. However, the review here of these studies reveals that not one of them demonstrates the greater effectiveness of either the emotional or logical communication. The three previous paragraphs reveal (1) that the result of the Hartmann study was in favor of the emotional appeal and was significant at the .50 level, (2) that the first result of the Menefee and Granneberg study was in favor of the emotional appeal and was significant at the .98 level, while their second result was in favor of the logical appeal and was significant at the .52 level, and (3) that the first result of the Knower study was in favor of the rational appeal and was significant at the .20 level, while the second result was in favor of the emotional appeal and was significant at the .95 level.

1.12 Janis and Feshbach (1953)

For the last 12 years the dominant research position in social psychology has held that an anxiety-inducing communication is less effective than a communication which contains the same information, but which does not induce anxiety. The argument tends to be centered around the idea of "defensive avoidance," a term intended to indicate that a subject will tend to avoid information which he receives in a communication which is threatening to him, with the result that, if the communication recommends a particular action, the subject will be less prone to engage in that action. In effect, the subject "defends" himself against the threat by not recalling and utilizing the information and recommendations of the communication. The basic idea was advanced originally by Janis and Feshbach (1953).

This study is cited frequently in support of the position that emotional involvement tends to reduce the effectiveness of a communication. For example, Hovland, Janis, and Kelley (1953) state that the results of Janis and Feshbach support the"...generalization that a high degree of emotional tension tends to reduce the over-all effectiveness of a persuasive communication." Because of the general acceptance of the Janis-Feshbach hypothesis, and because of its historic importance, it deserves to be described in detail.

The Janis and Feshbach study (1953) utilized three versions of a 15-minute slide-tape presentation on dental hygiene so as to induce three levels of fear arousal in three groups, each composed of 50 high school freshmen, all about 15 years old. They were equated with regard to IQ and sex, with about an equal number of boys and girls in each group. The strong fear-arousing communication emphasized the painful consequences of tooth decay, diseased gums, and other dangers that can result from improper dental hygiene, and utilized a personal presentation explicitly directed to the audience. The moderate fear-arousing communication presented dangers in a milder, more factual manner, and used an impersonal presentation. The minimal fear-arousing communication rarely alluded to the consequences of tooth neglect, and used an impersonal, factual presentation. All three versions contained the same essential information and the same recommendations concerning oral hygiene practices. An equated control group of 50 subjects was exposed to an irrelevant, though similar, communication on the structure and functioning of the human eye.

Janis and Feshbach tested first for differences between the groups with respect to affective reaction. There were actually two such tests. The first test was based on verbal responses obtained once - immediately after exposure to the communication. These responses were to three questions intended to measure feelings which the student experienced while the communication was being presented.

Although Janis and Feshbach (1953, p. 81) state that the students' responses to these three items "... indicate that the fear stimuli were successful in arousing affective reactions," the data which they present in support of this statement are insufficient to test this claim. In order to support their claim, it would be necessary to show that the experimental data differed from control data. The latter data were not obtained, since it would not have made sense to ask the control students immediately after exposure to a communication how worried they were about their mouth condition during exposure to a communication on the structure and functioning of the human eye. Hence, with respect to this measure, no statement can be made regarding the success of the fear stimuli in arousing affective reactions.

Apparently what Janis and Feshbach had intended to state was that the responses to the three questionnaire items indicated differences in arousal among the three experimental groups, which is the only matter of interest here, anyway. Unfortunately, they do not give the probability of the null hypothesis that the three groups are from the same population, and do not give any indication of having computed this probability. Thus, it becomes necessary to compute it here in order to determine the likelihood that Janis and Feshbach achieved an experimental effect.

In their Table 2, Janis and Feshbach (1953) show for each item separately the number of students in each group who gave a worried response. It is unfortunate that this is the only information given, for it precludes an analysis of variance (or covariance), which would have been proper here, and which Janis and Feshbach could have performed on the basis of the original anxiety scores, which covered a 16-point range. The reader who has access only to the published article is forced to utilize the more impoverished data presented in Table 2.

By averaging over the items, one obtains a more reliable measure on which to determine the extent to which the groups differ among themselves with respect to this 3-item measure of experimentally-induced anxiety. The Strong group shows the greatest number of worriers, the Moderate next, and the Minimal the least, and it is this trend, apparently, which influences Janis and Feshbach (1953, page 82) to conclude that "...the foregoing evidence indicates that after exposure to the communications, the Strong group felt more worried about the condition of their teeth than did the other two groups; the Moderate group, in turn, tended to feel more worried than the Minimal group."

It is not at all clear why the authors chose to divulge in another article published at a later time and in another journal (Janis and Feshbach, 1954) the fact that they obtained one week before the treatments a measure of anxiety, and that even before the treatment the Strong group contained the highest proportion of students whom the authors identified as scoring higher in chronic level of anxiety with respect to physical health; conversely the Minimal group contained the lowest proportion.

Thus, the analysis by Janis and Feshbach (1953) of the three questionnaire items in their Table 2 fails in two ways as being appropriate to a determination of whether the three experimental groups differed significantly with respect to post-communication anxiety - that is, whether Janis and Feshbach achieved a significant differential treatment effect. First, they made no overall analysis of all three items and all three groups. And second, they failed to control for pre-communication anxiety even though they discussed, and presented evidence of, the influence of the pre-communication anxiety scores on the post-communication anxiety scores (Janis and Feshbach, 1954, page 161). The analysis here will attempt to rectify both these limitations.

In determining the probability that the three groups are from the same population, the analysis here will be based on a comparison of obtained and expected frequencies in a 2x3 table (average over the three items of number of students indicating anxiety or no anxiety; Strong Moderate, or Minimal), with cell frequencies limited by marginal totals. Under this restriction, the probability of the null hypothesis that the groups are from the same population is equal to the probability of the null hypothesis that there is no interaction in the 2-way table (Kullback, 1959). The observed frequencies (cells and margins) being given directly by Table 2 in Janis and Feshbach (1953), it remains only to determine the expected frequencies.

The expected cell frequencies are taken as directly proportional to the marginal frequencies under the hypothesis that the three groups do not differ with respect to the criterion measure (in this case, whether the student is identified as high anxiety after the communication). However, this is not the hypothesis which should be utilized here, for it is known that the groups differed on a related measure before the communication. Thus, the null hypothesis which is appropriate is H_o, that the groups do not differ with respect to the extent to which they have been influenced by the communication. Under this hypothesis, we would expect the three groups to differ among themselves after the communication only to the same extent as they differed among themselves before the communication.

Since differences after the communication are manifested in the observed frequencies, differences before the communication need to be incorporated in the expected frequencies. This incorporation will now be described.

Based upon the marginal frequencies given by the data of Table 2 in Janis and Feshbach (1953), the expected number of worriers (or high anxiety students) is 23.11 in each group. This frequency needs to be adjusted (e.g., increased for the Strong group) in proportion to the differences among the groups with respect to pre-communication anxiety. For example, before the communication the Strong group contained 1.100 times as many high anxiety students as the Minimal group (Janis and Feshbach, 1954), and this same factor needs to exist for the expected frequencies.

The determination of the expected frequencies is complicated by the fact that, while the two articles (Janis and Feshbach, 1953; Janis and Feshbach, 1954) involve the same data, the authors have chosen to delete some of the students from the first article, and one of the groups (Moderate) from the second article.

The authors give as their justification for the deletion of some of the students in each group the improvement of the matching of these three groups (Janis and Feshbach, 1954, page 159). The matching variable(s) is not specified, but presumably it is a relevant variable; if so, it is hard to imagine a matching variable which is relevant in the first article but not in the second. It would seem that when it is sufficiently important to delete subjects in order to improve matching, it would be sufficiently important to specify the matching variable. Even more surprising is the fact that in the article (Janis and Feshbach, 1953) in which the reduced data are used, no mention is made of the fact that the data reported constitutes only a selected portion of the data collected; that is, no mention is made of deletion, much less its basis. In fact, the contrary statement is made: "Altogether there were 200 students in the experiment, with 50 in each group." (Janis and Feshbach, 1953, page 80), a statement flatly contradicted in the second article (Janis and Feshbach, 1954, page 159).

The authors give as their justification for the deletion of the Moderate group from the second article the desire to assess the interaction between pre-communication anxiety and communication threat (Janis and Feshbach, 1954, page 157). However, they fail to indicate how the inclusion of this portion of the collected data (the Moderate data) would have in any way interfered with their assessment of this interaction.

The anomalous procedures described in the preceding three paragraphs necessitate two assumptions in the specification of the expected frequencies. First, in the absence of any indication by Janis and Feshbach (1953; 1954) to the contrary, it is assumed that the proportion of pre-communication high anxiety students is the same in both the deleted and undeleted samples. Second, in the absence of any indication by Janis and Feshbach (1953; 1954) to the contrary, it is assumed that the Moderate group bears the same relation to the Strong and Minimal groups with respect to the proportion of students identified as high anxiety on the pre-communication measure as the Moderate group bears to the Strong and Minimal groups with respect to the proportion of students identified as high anxiety on the post-communication measure.

These assumptions lead directly to expected frequencies of high anxiety students (worriers) which have the following relations: 1.100 for the Strong group, 1.037 for the Moderate group, and 1.000 for the Minimal group. Utilizing these values to adjust the previously-given average expected value of 23.11, given the marginal restriction, the following expected values are obtained: 24.31 for the Strong group, 22.92 for the Moderate group, and 22.10 for the Minimal group. The six discrepancies between observed and expected values yield a chi square of 4.88 with 2 df, which is associated with a probability level of .09. Because of the two assumptions involved in the determination of this value, the value itself is, of course, an estimate.

Fortunately no such assumptions need be made with regard to the second Janis-Feshbach measure of affective arousal by the communications. This second measure was obtained both before and after the communication from each of the 50 students in each of the three experimental groups. (These 150 students are the only ones considered hereafter in the remainder of this review of the Janis-Feshbach study). This procedure allows the three groups to be compared on change measures, or better yet, on change measures adjusted to be independent of initial measures. In this latter case, the initial differences in chronic anxiety reported by Janis and Feshbach (1954) become unimportant.

The second test employed by Janis and Feshbach to determine whether the three experimental conditions differentially aroused affective reactions was a test based on verbal responses to two questions. These two questions were also intended to indicate the amount of anxiety experienced by the student about diseased gums and decayed teeth. However, since these two questions made no reference to the illustrated talk, it was feasible to administer them to each of the groups in the pre-communication and immediate post-communication questionnaires. Janis and Feshbach give the before and after percentage of students in each group who reported feelings of relatively high disturbance on both items. (These two items were considered by Janis and Feshbach as a pair, whereas the three items above were considered separately; no explanation for this procedural difference is offered.) The authors then compute the difference between pairs of groups with respect to change (from before-to-after, hereafter referred to simply as before-after) in anxiety response.

Again with respect to this second measure, Janis and Feshbach give no evidence of whether the three groups differ significantly among themselves with respect to affective reaction. Again what is needed is the probability of a single result summarizing the overall extent to which the groups differ among themselves. Until it has been shown that the groups differ significantly in an overall sense, it is simply not appropriate to proceed to test differences between pairs of groups. To do so is not only inappropriate but also misleading for the uncritical reader, who is likely to assume that differences between groups are indicative of the overall discrepancy among the groups. Such an interpretation is particularly likely in view of the use by Janis and Feshbach of the terms "Strong," "Moderate," and "Minimal" throughout their article without first having shown that the three groups differ significantly in affective reaction.

In view of the importance attributed by Janis and Feshbach to the differential effect of the three communications on affective reaction, and in view of the absence in the published account (Janis and Feshbach, 1953) of an overall statistical test of this effect, an attempt will be made at this time to determine whether the three experimental groups differ significantly among themselves.

In the questionaire as presented to the students, the two items allowed for a 5-point response, resulting in a 10-point spread in scores, which assuming normality and homogeneity, is sufficient for an analysis of covariance among groups, with the pre-communication measure as the covariate. Such an analysis is precluded, however, because the authors, for reasons not offered in the publication, chose to dichotomize the 10-point scale.

The only statistical information offered the reader is contained in Table 3 (p. 81), which contains only three items: (1) proportion of before-worriers in each group, (2) proportion of after-worriers in each group, and (3) critical ratios for the differences between pairs of groups (e. g., Strong vs. Moderate) with respect to before-after group change in anxiety response.

Any test to determine whether the groups differ significantly among themselves requires a knowledge not only of group change in anxiety but also of the proportions in the four subgroups which comprise each group. These four subgroups are (a) unworried-before, worried-after, (b) worried-before, worried-after, (c) unworried before, unworried-after, and (d) worried-before, unworried-after. This critical information, though not supplied by Janis and Feshbach, can be determined on the basis of the information given in their Table 3. The procedure involved in such a determination will be described at this time.

The procedure will involve two steps. First, a determination will be made of the four subgroups for pairs of groups combined (e. g., Strong plus Moderate). Secondly, from this information, a determination will be made of the proportions in the four subgroups for each group.

Janis and Feshbach calculated the critical ratios in Table 3 by way of the equation,

$$\mathrm{CR} = \frac{\mathrm{diff(a-d)}}{\mathrm{SE}_{\mathrm{diff(a-d)}}} , \qquad (1.\ 1)$$

where a is the proportion of a group (e. g., Strong) showing increased anxiety, d is the proportion showing decreased anxiety, diff(a-d) is the difference between two groups (e. g., Strong vs. Moderate) with respect to the difference in proportions of the two types of changers. Since diff(a-d) and CR are given in Table 3, $\mathrm{SE}_{\mathrm{diff(a-d)}}$ can be calculated. It is understood, of course, that this standard error is an estimate, as is almost always the case in practice, and as will be the case throughout this report. With this understanding, sampling statistics will not be explicitly designated as estimates.

The standard error of the difference between two independent groups (e. g., Strong vs. Moderate) with respect to the difference between the two types of changers was calculated by Janis and Feshbach from the equation,

$$\mathrm{SE}^2_{\mathrm{diff(a-d)}} = \mathrm{SE}^2_{\mathrm{(a-d)}_1} + \mathrm{SE}^2_{\mathrm{(a-d)}_2} \qquad (1.\ 2)$$

where $SE_{(a-d)_1}$ is the standard error of the difference in proportions of the two types of changers in sample 1 (e.g., Strong group).

In a footnote (p. 81), Janis and Feshbach (1953) indicate that they calculated $SE^2_{(a-d)_1}$ using the equation given in Hovland, Lumsdaine, and Sheffield (1949), namely,

$$SE^2_{(a-d)_1} = \left[\frac{A_c + D_c}{N_c} - \left(\frac{A_c - D_c}{N_c}\right)^2\right]\left(\frac{1}{N_1}\right), \tag{1.3}$$

where the subscript c stands for the combination of two experimental groups, where $N_c = 100$, where $N_1 = N = 50$, and where A_c and D_c are change frequencies, as indicated in Table 1.1 below. Here, and throughout this report, capital letters represent frequencies, and lower case letters represent the corresponding proportions.

TABLE 1.1

FOURFOLD FREQUENCY SYMBOLS DESIGNATING WORRIED AND UNWORRIED STUDENTS IN THE JANIS-FESHBACH STUDY (1953)

		Before	
		Unworried	Worried
After	Worried	A	B
	Unworried	C	D

The final objective here is to determine the value of A, B, C, and D for each experimental group, and this determination can be made if A_c, B_c, C_c, and D_c are known for combined pairs of experimental groups. These latter values can be derived, but before doing so, it should be noted that $a = A/N = A/50$, that $d = D/N = D/50$, and that $a - d = (A+B)/N - (B+D)/N = (A-D)/N$.

As mentioned previously, Table 3 in Janis and Feshbach (1953) gives the values of $(A+B)$ and $(B+D)$ for each experimental group. By direct addition of these values for two groups, the combined values of (A_c+B_c) and (B_c+D_c) are readily determined. By subtracting the latter from the former, the value of (A_c-D_c) is obtained. Substituting this value in equation (1.3), the value of whose left member is already known, the value of (A_c+D_c) is obtained. Knowing both (A_c+D_c) and (A_c-D_c) leads directly to a knowledge of A_c and D_c. These cell entries, when subtracted from the marginal totals, lead to B_c and C_c.

Table 1.2 gives the values of A_c, B_c, C_c, and D_c for the three possible pairs of experimental groups. This, then, completes the first step toward the determination of the cell frequencies for each experimental group.

TABLE 1.2

FREQUENCIES OF WORRIED AND UNWORRIED STUDENTS FOR PAIRS OF EXPERIMENTAL GROUPS IN THE JANIS-FESHBACH STUDY (1953)

		Before		
	After	Unworried	Worried	
Strong + Moderate	Worried	40	23	63
	Unworried	31	6	37
		71	29	

		Before		
	After	Unworried	Worried	
Strong + Minimal	Worried	38	23	61
	Unworried	34	5	39
		72	28	

		Before		
	After	Unworried	Worried	
Moderate + Minimal	Worried	34	14	48
	Unworried	43	9	52
		77	23	

The second step involves the determination of the cell frequencies of each experimental group by a procedure which utilizes the information in Table 1.2 above and in Table 3 of Janis and Feshbach (1953).

Table 3 in Janis and Feshbach gives the marginal totals of (A+B) and (B+D), where the unsubscripted letters are used here to represent values for an experimental group. Knowing that there are 50 subjects in an experimental group, the other two marginal totals, (A+C) and (C+D), are readily determined.

The 8 cell entries in the fourfold tables for two experimental groups (e.g., Strong and Moderate) involve a single degree of freedom, for the specification of a value in one table prescribes the corresponding value in the other table, for their sum is known and is given in Table 1.2 of this report. This fact allows a unique determination of the cell entries for each experimental group. For instance, there is only one possible value of A in the Strong table which restricts A in the Moderate table and A in the Minimal table to values which sum to 34 (A_c in the Moderate + Minimal table).

The cell entries of each experimental group are given in Table 1.3 of this report. The data in Table 1.3 here lead to the results shown in Table 3 of the Janis and Feshbach article, and, in fact, necessarily constitute the data used by Janis and Feshbach in determining the results shown in their Table 3.

Having determined the cell frequencies for each of the experimental groups, we can return to the initial question of whether these groups differ significantly among themselves with respect to affective reaction. As stated previously, the recommended procedure would have been an analysis of covariance, or an analysis of variance in which the change scores were transformed in some way to make the transformed scores independent of the original scores (e.g., Levonian, 1963a). Either

TABLE 1.3

FREQUENCIES OF WORRIED AND UNWORRIED STUDENTS FOR EACH EXPERIMENTAL GROUP IN THE JANIS-FESHBACH STUDY (1953)

		Before		
	After	Unworried	Worried	
Strong	Worried	22	16	38
	Unworried	11	1	12
		33	17	
		Before		
	After	Unworried	Worried	
Moderate	Worried	18	7	25
	Unworried	20	5	25
		38	12	
		Before		
	After	Unworried	Worried	
Minimal	Worried	16	7	23
	Unworried	23	4	27
		39	11	

type of analysis could have been performed by Janis and Feshbach from their original data, but neither procedure can be performed at this time by a person who has access only to the published article (Janis and Feshbach, 1953). The published data are restricted to frequencies, and these must form the basis for the determination of the probability of the null hypothesis that the three experimental groups are from the same population. This determination will be made here by two independent methods, the first leading only to an approximation of the desired probability.

The first method, which has a limitation to be discussed later, involves combining three probabilities, one for each experimental group, where each probability is determined under the null hypothesis that each group is from the same population. For this purpose, we need an estimate of the standard error of the difference in proportion of the two types of changers in a sample of size N, as given by Equation 1.3.

It is recalled that Janis and Feshbach (1953) calculated three estimates of $SE^2_{(a-d)}$ by employing successively three sets of values for the expression in brackets in Equation 1.3. The first set was obtained by combining the Strong and Moderate groups, the second set by combining the Strong and Minimal groups, and the third set by combining the Moderate and Minimal groups. If one truly assumes that the three groups are from the same population, then it would seem that the combination of all three groups would yield a better estimate of the sampling error for this population. Unfortunately, Janis and Feshbach gave no rationale for the procedure they employed.

In any case, since our null hypothesis is that the three experimental groups are from a single population, values for the bracketed term in Equation 1.3 will be based on the pooled data of the three groups. The pooled values are shown in the Total column of Table 1.4, which presents the frequencies of Table 1.3 in a more convenient format.

TABLE 1.4

FREQUENCIES OF CHANGES IN REPORTED DENTAL ANXIETY FOR EACH EXPERIMENTAL GROUP IN THE JANIS-FESHBACH STUDY (1953)

Type of Change	Strong	Moderate	Minimal	Total
Increased Anxiety (A)	22	18	16	56
Decreased Anxiety (D)	1	5	4	10
No Change (B+C)	27	27	30	84
(A-D)	21	13	12	46
(a-d)	.42	.26	.24	.3067

When the A and D values in the right-most column of Table 1.4 are inserted into Equation 1.3, with $N_c = 150$ and $N = 50$, the $SE_{(a-d)}$ which emerges has the value .0832. The last row of Table 1.4 gives for each group the observed difference in proportion of the two types of changers. This difference for the population is best estimated from the pooled differences for the three experimental groups, and this estimate of .3067 is also shown in Table 1.4.

When the difference between .3067 and the (a-d) value for each group is tested against the standard error of .0832, the three t values which emerge are 1.36 (Strong), .56 (Moderate), and .80 (Minimal). These three values need to be combined in such a manner as to result in a single probability of our null hypothesis.

Fisher (1954) has described how a probability may be converted to a chi square with 2 degrees of freedom, and how several such chi squares, if independent, may be added to yield a chi square of the combined probability (this procedure will be discussed in detail later). Thus, by utilization of the probabilities of the three t ratios given in the above paragraph with the table supplied by Gordon, Loveland, and Cureton (1952), or with the equation,

$$\chi^2 = -2 \sum_{i=1}^{k} \log_e p_i, \tag{1.4}$$

the combined χ^2 of 10.50 with 6 df shows a probability of .35. This, then, is the single probability value which was desired. It represents the probability of the null hypothesis that the three groups are from a population with an (a-d) value of .3067.

As was stated previously, this null hypothesis probability of .35 is only an estimate, a reflection of the fact that only two of the three t ratios are independent. Thus, if the values in Table 1.4 for one of the groups were not known, these values could be inferred from the remaining three columns.

A second method, which is free of the limitation just mentioned, will be employed now in the determination of the probability of the null hypothesis that the three experimental groups are from the same population. As mentioned previously, the proper type of analysis would have been analysis of variance or covariance. Unfortunately this type of analysis is no longer possible from the categorical data offered the reader by Janis and Feshbach (1953). The question which now arises is whether a form of variance analysis can be applied to multinomial data, data which consist solely of unordered categories. This should be possible, for such data may be expressed in informational terms, and, as McGill (1955) points out, the structure of information analysis is analogous to variance analysis. The objective in each type of analysis is to partition some quantity - information in one case, variance in the other - into components representing main effects and interactions. Kullback (1959) supplies an exceedingly detailed account of the relation between information theory and classical statistics.

The significance of the differences among the category frequencies of the Strong, Moderate, and Minimal groups, as shown in Table 1.4, may be computed using the information statistic.

Let I represent the mean information in the row categories about samples from the same population. For I, Kullback (1959) gives the expression,

$$I = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij} \log \frac{x_{ij}}{N_j p_i} , \tag{1.5}$$

where the logarithm (as throughout this report) is to the Napierian base, the i designate rows, the j designate columns, the x_{ij} are the cell frequencies, the N_j are the sums of the frequencies in each column, and the p_i are the parametric probabilities, that is, the population probability for each row. The hypothesis H_o, that the samples are from the same population, need not, and often does not, specify the p_i, i=1, 2,... r.

It may be appropriate at this time to point out two of the characteristics associated with the information statistic of Equation 1. 5. First, it can be shown that that statistic is asymptotically distributed as chi square with c(r-1) degrees of freedom (Kullback, 1959).

Second, the information represented by Equation 1. 5 can be analyzed into two additive components: I_w due to the variability within samples, and I_b due to the deviations between the p_i and their best unbiased estimates from the pooled samples. Each of these components is distributed as chi square, the first with (r-1)(c-1) df, the second with (r-1) df. Each of these components tests a different aspect of the null hypothesis, H_o, that the samples are from the same population. As Kullback (1959) points out, H_o is the intersection of two hypotheses: (1) that the samples are homogeneous, and (2) that the homogeneous samples are from the population $(p) = (p_1, p_2, \ldots p_r)$. I_w tests the first hypothesis, while I_b tests the second. If the null hypothesis does not specify the p_i, as is usually the case, as in the Janis-Feshbach study, the p_i are estimated from the pooled samples, and I_b vanishes. In this case, which is the case to be considered at this time, I_w is the only component in I. Thus, the equations for I_w and I_b will not be given here, and the value of I_w may be taken as that given by Equation 1. 5. When the p_i are estimated from the pooled samples, an approximation to the I_w chi square may be obtained from the standard Pearson chi square technique for testing the independence of rows and columns in a contingency table, with estimated values based on marginal totals (Kullback, 1959); in this case the information statistic and the Pearson chi square are equivalent (Kendall and Stuart, 1961, page 421).

We turn now to the application of the information statistic to the Janis-Feshbach data. First, before determining the probability of the overall H_o, let us compare the behavior of the information statistic relative to the (a-d) statistic employed by Janis and Feshbach. Consider first the data given in Table 1. 4 for the Strong and Moderate groups.

Utilizing these six values with Equation 1.5, we obtain a chi square of 3.31, which for 2 df yields a probability of .19. Similarly, the six values for the Strong and Minimal groups result in a chi square of 3.04 having a probability of .22. Finally, the data from Table 1.4 for the Moderate and Minimal groups yield a chi square of .39, which has a probability of .83. These three probabilities - .19, .22, and .83 - are similar to, but not identical with, the three corresponding values of .18, .12, and .86 given by Janis and Feshbach and computed from Equations 1.1, 1.2, and 1.3 given in this report.

The discrepancy between a pair of corresponding probabilities is due to the fact that each is the probability of a different null hypothesis. The probability computed here is the probability that two groups (e.g., Strong and Moderate) do not differ with respect to the proportion of students in three categories (A, B+C, and D), whereas the probability computed by Janis and Feshbach is the probability that two groups do not differ with respect to a-d, the difference in proportions in the A and D categories. Which constitutes the more appropriate null hypothesis: (1) that two groups do not differ over all categories, or (2) that two groups do not differ in the difference between two categories?

There is one characteristic of Equation 1.1 which, in some applications, might be considered a limitation in the sense that it requires a decision on the part of the investigator before the data are inspected. That equation, in assessing the difference between two categories, must actually recognize three, but only three, categories: the first category of interest, the second category of interest, and a third category. In many applications this third category is actually composed of several categories, which are pooled only because such a procedure is required by Equation 1.1.

For instance, Janis and Feshbach pool the frequencies in the B and C subgroups (see Tables 1.1 and 1.4 of this report). The logic of this is not made clear by Janis and Feshbach, but it would seem that students worried-before and worried-after (B) may be as different from students unworried-before and unworried-after (C) as the two changer types (A and D) are from each other. Combining categories generally reduces the information contained in data, and may lead to an interpretation of the pooled results which would not seem to be an appropriate interpretation of the unpooled results.

It should be noted that the necessity of pooling all proportions in a sample other than a and d is not an inherent limitation of Equation 1. 1. That equation is appropriate for assessing the difference between two samples with respect to (a-d), and not with respect to any other differences which may exist between the samples. This fact, however, imposes a requirement on the utilization of Equation 1. 1, namely, that the two change categories (those corresponding to a and d) be specified, either on a procedural or theoretical basis, before the data are inspected. Janis and Feshbach (1953) give no information on this point, except to say that a worrier was defined as one who reported a relatively high disturbance in response to both questions. Thus, a student who reported "slightly worried" is counted as unworried in Tables 1. 3 and 1. 4, whereas a student who reported "somewhat worried" is counted as worried.[1] Presumably this decision was made before the data were inspected.

Having made explicit these characteristics of Equation 1. 1, we can return to the question of the merit of comparing two samples from a multinomial population by (1) comparing on all categories with Equation 1. 5, or (2) comparing on two selected categories with Equation 1. 1.

It seems to me that the information statistic is much to be preferred. First, it considers frequencies (or if you wish, proportions) in each category, not simply the difference between two categories, and by so doing, does not lead to the anomalous results which occur with Equation 1. 1, for extreme values of the a's and d's. Second, it imposes no restrictions on the number of categories; yet in the degenerate case of the binomial population, the information statistic yields precisely the same probability as the classical statistical procedure involving the error variance given by

---

[1] These alternatives used by Janis and Feshbach (1953) were not given by them in their article, so it was necessary to refer to the more complete report by Goldstein (1957), who used these same items, for the wording of the alternatives.

$$SE^2_{diff(p)} = pq \left(\frac{1}{N_1} + \frac{1}{N_2}\right). \tag{1.6}$$

Third, it allows the computation of the probabilities of three null hypotheses: (1) that the two samples are homogeneous, (2) that the two samples are from the same unspecified population, and (3) that the two samples are from the same specified population (one whose parameters are specified). All in all, it would seem that the information statistic is to be preferred over Equation 1. 1 when two groups are being compared.

The elegance and utility of the information statistic emerges even more clearly when it is recognized that the information measure is appropriate not only for two samples but also for k samples, not only for 2-way analysis but also for n-way analysis, not only for multinomial populations but also for continuous populations, and not only for univariate problems but also for multivariate problems. As such, it articulates readily with analysis of variance (Garner and McGill, 1956; McGill, 1954).

Finally, since information can always be written in the form n log n, or m log n, information measures can be tabled. Existing tables are to the base 2 (Air Force Cambridge Research Center, 1954; Attneave, 1959; Klemmer, 1955; Miller and Ross, 1954), to the base e (Kullback, 1959), or to the base 10 (Fisher, 1956; Miller and Ross, 1954). Sometimes it is more convenient to express informational measures in terms of -p log p, and tables for these exist, either to the base 2 Air Force Cambridge Research Center, 1954; Klemmer, 1955), or to the base e (Bartlett, 1952). And since information is additive, one needs to supplement these tables with nothing more than a calculating machine.

Equation 1. 1 and the Equation 1. 5 are similar in one negative sense: neither is appropriate for determining the probability that the Strong, Moderate, and Minimal groups are from the same population (which, if the parameters of that population are unspecified and are taken to be the unbiased estimates from the pooled samples, is the same as determining the probability that the three groups are homogeneous) by a procedure which involves two samples at a time. It is this procedure which Janis and Feshbach have used in their Table 3, and which has led to this long discussion.

Why, then, was this same procedure used a few pages back with Equation 1. 5? Simply to demonstrate that the application of Equation 1. 5 to samples taken two at a time leads to probabilities similar to those obtained by Janis and Feshbach with Equation 1. 1. With that fact firmly in mind, we are finally ready to apply Equation 1. 5 to the Strong, Moderate, and Minimal data in a single analysis.

When Equation 1. 5 is applied to the Janis-Feshbach data reproduced in Table 1. 4, the result is a chi square of 4. 30, which for 4 df yields a probability of . 36. Unfortunately, Janis and Feshbach (1953) fail to give a probability level at or beyond which they consider a result significant, and hence, it is not possible to state categorically that the differences among the Janis-Feshbach treatments are not significant by their criterion. However, by imposing the conventional . 05 level, the Janis-Feshbach data are insufficient to allow one to reject the null hypothesis that the three experimental groups are from the same population.

In all fairness to Janis and Feshbach, it should be mentioned that they never explicitly stated that they achieved a significant experimental effect, but the tenor of the article leads the reader to believe that this was the case. For instance, on page 80 of Janis and Feshbach (1953) one finds the following: "Evidence that the three forms of the illustrated talk differed with respect to the amount of emotional tension evoked during the communication is presented in Table 2." On page 82 they state that "In general, the foregoing evidence indicates that after exposure to the communications, the Strong goup felt more worried about the condition of their teeth than did the other two groups; the Moderate group, in turn, tended to feel more worried than the Minimal group." On page 90 they state that "...the fact remains that the 'unreduced' emotional tension was manifested immediately after the communication predominantly by the group exposed to the Strong appeal." And throughout the article, the authors refer to a particular appeal (e.g., the Strong appeal) as achieving, relatively, the effect (and affect) implied by the adjective.

The fact that this effect was significant only at the .36 level suggests two alternatives: (1) the authors had to claim a differential effect in order to support their defensive avoidance hypothesis, or (2) the authors erroneously believed that their treatments actually achieved a differential effect. It will be assumed here that the second alternative is correct, and on this basis, it will be assumed that Janis and Feshbach were victims of their own technical limitations.

While the Janis-Feshbach data on anxiety change are not inconsistent with the null hypothesis that the treatments are from a population with values a, d, and b+c the same as the pooled data, this is not the only possible null hypothesis.

There are, of course, an infinite number of null hypotheses which one would fail to reject given the Janis-Feshbach anxiety change data, but one interesting, non-arbitrary one is the null hypothesis that the three experimental groups are from a population whose parametric values for a, d, and b+c are given by the marginal frequencies of a fourfold table of the type shown in Table 1.1 for the pooled data of these three groups. Thus, for the three groups combined, 73.4% of the subjects were unworried before treatment and 57.2% were worried after treatment, yielding an a of .420. Similarly, the population value of d is taken as .114, and b+c as .466. Note that these expected values are not the same as the observed values previously used with Equation 1.5. When the current values of a, d, and b+c are used with Equation 1.5, the resulting chi square of 11.4, this time with 6 df, fails to reach the .05 level. Note also that the procedure just discussed is not simply the traditional application of Pearson chi square test to the fourfold frequency table. Thus, we are unable to reject the hypothesis that the three samples are from a population having 26.6% worriers before treatment and 57.2% worriers after treatment.

Inasmuch as the Janis and Feshbach study failed to demonstrate a differential treatment, it is unlikely that the three groups will differ significantly with respect to measures of communication effectiveness. Therefore, we can move quickly through the remainder of the article (Janis and Feshbach, 1953).

In order to determine whether the three groups differed significantly in the amount of information learned from the communication, a 23-item information scale was administered immediately after the communication. The authors give no information on the results except to indicate that the three groups did not differ significantly with respect to the information test scores.

In order to determine whether the three groups differed significantly in their appraisal of the communication, a 7-item appraisal scale was administered immediately after the communication. As usual, Janis and Feshbach (page 82) cite differences between selected pairs of groups on selected items, and conclude that "... it is apparent that the Strong group responded more favorably than the other two groups".

In evaluating this conclusion, it is important to recall again that when Janis and Feshbach identify the groups as Strong, Moderate, and Minimal, they are identifying the groups according to the amount of fear arousal; they are not using these adjectives in a non-quantitative sense, as one would in referring to these groups, possibly, as the X group, the Y group, and the Z group. Thus, when they state that "... students exposed to the Strong appeal were more likely than the others to give favorable appraisals concerning the interest value and the quality of the presentation," they imply that fear arousal and favorable appraisal are positively related.

The manner in which Janis and Feshbach generalize on the basis of their analysis of selected data leads me to suspect that they fail to recognize that occasional significant differences must necessarily arise between experimental groups, differences in no way related to the intended experimental effect, whether this effect was achieved or not. Speculation is one thing, but when one purports to utilize data in support of a generalization, one should use all the data gathered for the purpose.

Again, then, it becomes necessary to perform an overall analysis of the appraisal data. Unfortunately, the authors offer no analysis with respect to appraisal scores. For each item separately, the reader is offered only the number of students in each group who gave a favorable response. In the absence of the mean and standard deviation of the appraisal scores of each group, the best overall analysis is to compare the three groups with respect to the number of students giving favorable and unfavorable responses, averaged across items. When the three groups are compared on these frequencies (with the parametric values based on the pooled samples), the resultant chi square of 4.38 with 2 df is associated with a probability level of .12.

Janis and Feshbach supplemented their 7-item appraisal scale with two open-ended questions inviting criticisms of the illustrated talk. The authors categorized the student comments but published the frequency of comments for selected categories only. These frequencies are supplemented in the text with selected anecdotal material. This section can be ignored, not because it presents results based on essay responses, but because the authors fail to state the basis for their selection of the material to present to the reader. Such a specification is particularly important for the reader who knows that it is an unimaginative investigator indeed who would be unable to support practically any hypothesis from a selection of essay material supplied by 150 intellectually-active high school freshmen.

In order to determine whether the three groups differed significantly with respect to conformity to communication-recommended dental practices, a 5-item dental practices questionnaire was administered one week before and one week after the communication. The questions covered practices about which specific recommendations were made in the three forms of the communication. If on the second administration a student reported currently utilizing more of the five practices than he reported two weeks previously, he was listed under "increased conformity."

Janis and Feshbach analyzed these dental practices change data in the same way they analyzed the dental anxiety change data, but this time they give the reader the frequencies in the A, D, and B+C categories, resulting in the saving of labor which went into the determination of the frequencies in Tables 1.3 and 1.4. Why they elected not to present the earlier anxiety change data in this more informative form (the amount of tabular material is essentially the same) is not discussed by the authors. It would have been even simpler, and very much more informative, if the authors had given the A, B, C, and D frequencies for each group. In any case, the conformity change data given by Janis and Feshbach (1953, page 84) in their Table 6 is reproduced here in Table 1.5.

Again, without explicitly stating so, Janis and Feshbach imply that the three groups differ significantly with respect to this measure of conformity (reported change in dental practices in the communication-recommended direction), with increased conformity being negatively related to communication-induced anxiety. For instance, the authors state

TABLE 1.5

FREQUENCIES OF CHANGES IN REPORTED DENTAL PRACTICES
FOR EACH EXPERIMENTAL GROUP IN THE
JANIS-FESHBACH STUDY (1953)

Type of Change	Strong	Moderate	Minimal	Total
Increased Conformity (A)	14	22	25	61
Decreased Conformity (D)	10	11	7	28
No Change (B+C)	26	17	18	61
(A-D)	4	11	18	33
(a-d)	.08	.22	.36	.22

(page 84) that their data "... show a fairly consistent trend which suggests that as the amount of fear-arousing material is increased, conformity tends to decrease. In contrast to the marked increase in conformity produced by the Minimal appeal and the fairly sizable increase produced by the Moderate appeal, the Strong appeal failed to achieve any significant effect whatsoever." In another place (page 92) Janis and Feshbach state that "The evidence strongly suggests that as the amount of fear-arousing material is increased, conformity to recommended (protective) actions tends to decrease." Further, they state (page 85) that their results "... demonstrate that the Strong appeal was markedly less effective than the Minimal appeal..."

Such statements, however, would be markedly more effective if the authors presented evidence of the reliability of their results. Since they have not, it is necessary to do so here in order to assess the probability of the correctness of their conclusions.

Utilizing the procedure described previously, and applying Equation 1.5 to the data in Table 1.5, the mean information of the groups about the change categories (A, D, and B+C) yields a chi square of 6.63 with 4 df. This gives a probability of .16 for the null hypothesis that the three groups are from the same population. Again, in the absence of a statement by Janis and Feshbach regarding the level at or beyond which they consider results significant, we will utilize the conventional .05 level. Applying this level, we are forced to conclude that the Janis and Feshbach data given in Table 1.5 are not inconsistent with the null hypothesis that the three groups are from the same conformity population. Of course, this is the result one would expect on the basis of the relatively high probability that the three groups are from the same fear-arousal population.

Janis and Feshbach also used a second measure of conformity to communication-recommended behavior - in this case, going to the dentist. One week after the communication, the students were asked to indicate the approximate date on which they had last gone to a dentist. The numbers of students in each group who indicated having gone to the dentist during the week following the communication were as follows: 5 in the Strong group, 7 in the Moderate group, and 9 in the Minimal group. The chi square of the resulting 2x3 frequency table (dentist or no dentist; Strong, Moderate, or Minimal) is 1.33, and this is without a correction for the small N's in some cells. With 2 df, this chi square is significant at the .52 level. Undaunted, Janis and Feshbach cite these data as "...further evidence in support of the conclusion ...that the Strong appeal was markedly less effective than the Minimal appeal..."

Another measure of effectiveness was afforded by a 4-item scale administered one week before and one week after the communication. This scale measured the respondent's belief concerning the desirability of four communication-recommended characteristics of toothbrushes. The authors present no data, but simply state that, with regard to change in belief, there were no significant differences among the three experimental groups.

Finally, communication effectiveness was measured by the subject's resistance to counterpropaganda. The subject's attitude concerning the importance of using a toothbrush having the four communication-recommended characteristics was measured by a single item administered one week before and one week after the communication. However, at the second administration the item was preceded by an allegedly authoritative statement regarding toothbrush characteristics, a statement which contradicted communication statements in this regard.

The reader may wonder why it is that the respondent has a "belief" concerning the desirability of certain characteristics of toothbrushes, but an "attitude" concerning the importance of using a toothbrush with these characteristics. Janis and Feshbach do not discuss this distinction between belief and attitude, but we will continue to use the distinction they make. For convenience we will identify the attitude by its object, following conventional terminology (e. g., anti-Semitic attitude), leading to a toothbrush usage attitude.

Janis and Feshbach analyzed their toothbrush usage attitude change data in the same way that they had previously analyzed their dental anxiety change data and their dental practices change data. Their Table 7, giving the frequencies for each type of change, is reproduced here as Table 1. 6. In Table 1. 6 "increased conformity" means that the student's before answer to the item was not in the communication-recommended direction, whereas his after answer was.

TABLE 1. 6

FREQUENCIES OF CHANGE IN TOOTHBRUSH USAGE ATTITUDE FOR EACH EXPERIMENTAL GROUP IN THE JANIS-FESHBACH STUDY (1953)

Type of Change	Strong	Moderate	Minimal	Total
Increased Conformity (A)	19	21	27	67
Decreased Conformity (D)	15	14	7	36
No Change (B+C)	16	15	16	47
(A-D)	4	7	20	31
(a-d)	.08	.14	.40	.2067

Again, without explicitly stating so, Janis and Feshbach imply that the three groups differ significantly with respect to this measure of conformity (change in toothbrush usage attitude in the communication-recommended direction), with increased conformity being negatively related to communication-induced anxiety. They interpret their data as support for the conclusion that fear arousal is correlated negatively with resistance to counterpropaganda and with stable and persistent attitude changes. As usual, Janis and Feshbach do not make the assessment of whether the three groups differ significantly with respect to toothbrush usage attitude change, and, hence, it is necessary to perform the calculation here.

Utilizing the procedure described previously, and applying Equation 1.5 to the data in Table 1.6, the mean information of the three groups about the three possible toothbrush attitude usage change categories (A, D, and B+C) yields a chi square of 5.03, which with 4 df, is associated with a probability level of .29. As usual, this is the probability of the null hypothesis that the three samples are from a single population whose proportions a, d, and b+c are those corresponding to the Total column of Table 1.6 - that is, from a population whose parameters are estimated from the sample data. And, as usual, we are unable to reject this hypothesis at the .05 level.

In the final section of their Results section, Janis and Feshbach (1953) presented again selected categories of responses to an open-ended question which asked the student to give the reason for his answer (either true or false) to the one item intended to measure toothbrush usage attitude. Inasmuch as the authors fail to specify the basis on which they selected the categories to present to the reader, this final material is best ignored, for reasons already given.

This, then, brings us to the end of this review of the results of the Janis-Feshbach study, and we are in a position now to summarize the procedure and results.

The procedure involved the administration of three illustrated recorded talks to three groups of high school freshmen. The treatments were designated as Strong, Moderate, and Minimal, according to the experimenters' presumption before the talks were given of the degree of fear arousal which the talks would mediate. Measures on 8 variables were obtained at one or two of three points in time: (1) one week before the communication, (2) immediately after the communication, and (3) one week after the communication. All 8 measures were obtained from questionnaire responses. The points in time at which each measure was obtained, as well as the number of items associated with each measure, are shown in Table 1.7. The measures are listed in the order in which they were discussed in Janis and Feshbach (1953) and in this report.

TABLE 1.7

VARIABLES UTILIZED IN THE JANIS-FESHBACH STUDY (1953)

No.	Variable	No. of Items	1 Week Before	Just After	1 Week After	Probability of H_o
1	talk-related dental anxiety	3		x		.09[a]
2	talk-unrelated dental anxiety	2	x	x		.36
3	talk-related dental information	23		x		___[b]
4	appraisal of talk	7		x		.12
5	talk-recommended dental practices	5	x		x	.16
6	talk-recommended dental examination	1			x	.52
7	talk-recommended toothbrush belief	4	x		x	___[b]
8	talk-recommended toothbrush attitude	1	x		x	.29

---

[a] estimate

[b] reported by Janis and Feshbach as being not significant

Table 1. 7 indicates that Janis and Feshbach (1953) measured (a) anxiety aroused by the communication (Variables 1 and 2), (b) student appraisal of the communication (Variable 4), and (c) effectiveness of the communication (Variables 3, 5, 6, 7, and 8).

In addition to the 8 measures shown in Table 1. 7, Janis and Feshbach discussed and presented selected categories of responses to three open-ended questions (two supplementing Variable 4 and one supplementing Variable 8), but since the authors didn't specify the basis on which they edited the material to be presented to the reader, this material was not considered in this report.

The right-most column of Table 1. 7 lists the probability of the null hypothesis, H_o, that the three experimental groups are from a single population with parameters the same as the best unbiased estimates from the pooled samples. The absence of values for Variables 3 and 7 reflects the fact that no quantitative data were presented by Janis and Feshbach (1953) relating to these two variables. With respect to Variable 3 the authors state: "No significant differences were found among the three experimental groups with respect to information test scores." With respect to Variable 7, the authors state: "Among the three experimental groups, there were no significant differences with respect to net changes." With respect to the remaining six variables, the authors leave the impression that the three experimental groups differ significantly among themselves.

If the six probabilities in Table 1. 7 are independent, they may be combined into a single probability by the procedure already discussed in relation to Equation 1. 4. Before doing this, however, it is necessary to consider a point on which there is some apparent disagreement in the literature. The question is whether the probabilities used with Equation 1. 4 are restricted to one-tailed probabilities (Villars, 1951).

Most writers impose no such restriction (Johnson and Jackson, 1959; Wallis, 1942; Yule and Kendall, 1950), including the two persons who, independently, developed the method (Fisher, 1932; Pearson, 1933). However, some writers do (Gordon, Loveland, and Cureton, 1952; Guilford, 1956), apparently because of a failure to distinguish between the null hypothesis for which the combined probability is appropriate in general, and a more restricted interpretation of this null hypothesis. The examples given by Gordon, Loveland, and Cureton, and by Guilford, lead the reader to assume that these authors are considering only the more restricted interpretation, one for which one-tailed probabilities are required. However, it does not appear that this is a general requirement, and it is to this matter that we now turn.

The only general requirement for using Equation 1.4 is that each probability be from a rectangular distribution, $0 \leq p \leq 1$ (Kendall, 1946; Wallis, 1942). Thus, probability is considered a random variable, and the probability of obtaining a probability value of, say, .30 or less is .30. Such a rectangular distribution results from sampling a population under the null hypothesis appropriate to that population. For instance, given a population with zero mean and unit variance, the random selection of deviates and their conversion to probabilities will result in a rectangular probability distribution. There is no requirement that each probability to be combined pertain to the same null hypothesis, or that each probability is one-tailed or two-tailed. The probability associated with the chi square obtained from Equation 1.4 pertains only to the null hypothesis that each probability is from a rectangular distribution.

Now, it so happens that a common application of Equation 1.4 is the determination of the probability of one and the same null hypothesis tested repeatedly, but independently, say by different investigators or by the same investigator with repeated samples. It may be impossible or inappropriate to combine these data. In this common case, if the probability yielded by Equation 1.4 results in a rejection of the combined null hypothesis (that each probability is from a rectangular distribution), such a rejection is tantamount to a rejection of the null hypothesis for which each of the probabilities represents a Type I error. If, in addition, this initial hypothesis is one-sided, then a one-tailed probability yields a rectangular distribution.

It is this last case only which appears to have been considered by those writers who specify as a requirement the use of one-tailed probabilities with Equation 1. 4. However, it will be assumed here that it is the more general null hypothesis which is tested by the probability given in Equation 1. 4, and under this assumption it is not necessary to place restrictions on (1) the type of continuous distribution on which any of the probabilities is based (one might be based on an F distribution, another on the t distribution, still another on a chi square distribution), (2) the nature of the hypothesis (whether one-sided or two-sided) to which the probabilities pertain (Wallis, 1942), or (3) the nature of the data from which the probabilities are derived (Fisher, 1954).

Before returning to the Janis and Feshbach results, it should be noted that the use of Equation 1. 4 with probabilities one or more of which are derived from a discrete distribution (or a continuous distribution used to approximate a discrete distribution, as is the case with the six probabilities in Table 1. 7) will generally result in a composite probability which is biased upward (Wallis, 1942). However, if the number of steps in the distributions (essentially equivalent to the smallest marginal frequency for each of the six groups of data represented in Table 1. 7) is not small, which is the case here, the error is slight and may be ignored.

Applying Equation 1. 4 to the six probabilities of Table 1. 7, a combined probability value of . 10 is obtained. Thus, one is unable to reject at the . 05 level the null hypothesis that each probability is from a rectangular distribution. The implication of this for the Janis-Feshbach study is readily seen upon review of the procedure employed in deriving the six probabilities.

Each entry in the right-most column of Table 1.7 gives the probability of the null hypothesis that the three experimental groups are from a population with parameters the same as the best unbiased estimates from the pooled data (the first probability differs only in that the estimated parameters were adjusted slightly to reflect pre-communication differences among the groups with respect to a relevant covariate). As discussed previously, when the population parameters are unspecified, as was the case here, each probability also applies to the null hypothesis that the groups are homogeneous. There are, of course, six such null hypotheses, and we are unable to reject them as a group.

Therefore, the Janis-Feshbach data fail to yield sufficient evidence to reject the null hypothesis, H_o, that the three groups are homogeneous with respect to a population of variables of which the six in Table 1.7 are representative. Other variables might be shoe size, visual acuity of the left eye, and number of right turns made by the student in walking to school.

The preceding paragraph carries with it the qualifications already stated previously in this review: (1) one of the six probabilities is an estimate, and (2) the probability of .10 is based only on six of the eight measures used by Janis and Feshbach. Both of these qualifications probably lead to a lower probability for H_o, stated in the previous paragraph, than is in fact the case. If the one estimated probability is deleted, the combined probability based on the remaining five probabilities is .19, rather than .10.

Regarding the second qualification, the reader is left with the impression that the two probabilities for Variables 3 and 7 are greater than the six probabilities for the other six variables. This impression is based primarily on the fact that it is only with respect to the results of Variables 3 and 7 that Janis and Feshbach use the term "not significant" (page 82 for Variable 3 and page 85 for Variable 7), whereas with respect to each of the results of the remaining variables, Janis and Feshbach use the term "significant" and/or "reliable" (page 81 for Variable 1, page 81 for Variable 2, page 82 for Variable 4, page 84 for Variable 5, page 85 for Variable 6, and page 86 for Variable 8). If it is true that the null probabilities for Variables 3 and 7 are greater than the remaining six variables, one would expect, in general, to find the combined probability of all eight probabilities to be greater than .10. Thus, the value of .10 is an estimate but one which with respect to H_o, that the three experimental groups are homogeneous with respect to a population of variables of which the six in Table 1.7 are representative, is most likely biased in a direction which would lead to a Type II error.

It should be recalled that the interpretation given here is based on the assumption that the six probabilities are independent. If the raw data were available, this assumption could be tested by computing correlations between all possible pairs of the six variables; if these correlations were no larger than would be expected on the null hypothesis of zero correlations between each pair of variables, the six probabilities could be assumed to be independent. This procedure is not available to the reader who has access only to the published article (Janis and Feshbach, 1953). But even if it were true that there is a correlation between the six variables because, say, the Strong (Minimal) group scores higher (lower) on both anxiety and appraisal, as Janis and Feshbach claim, then such correlations would lead to a spuriously low probability of H_o - that is, the value of .10 would be biased downwards. This may be seen by eliminating, one at a time, each probability in Table 1.7 from Equation 1.4, and noting the effect on the composite probability. The value of .10 is increased when, successively, the probabilities associated with Variables 1, 4, and 5 are not included, and the value of .10 remains essentially unchanged when the probability associated with Variable 8 is not included. In effect, then, only two of the six variables, by virtue of their inclusion, tend to lead to a value as high as .10. Thus, the majority of variables contribute toward lowering the probability of H_o. If, now, the six variables were correlated because of the experimental treatments as Janis and Feshbach imply, then the effect of one would tend to be duplicated by the effects of the others - that is, all (or at least a majority) of the variables would tend to move the probability of H_o in the same direction. Since it has been shown that the majority of the variables tend to move the probability of H_o downward, we may assume that any majority effect which may exist due to a correlation among variables, as implied by Janis and Feshbach, would tend to lead to a probability of H_o which is lower than it would be if there were no such correlation. In summary, then, of this paragraph, while no test can be made of the assumption of independence of the probabilities which were combined, the existence of any dependence of the type implied by Janis and Feshbach is most likely to make .10 a lower value than is actually the case.

All in all, it seems probable that the value of .10 is, if anything, lower than the true probability of H_o, that the three experimental groups are homogeneous with respect to a population of variables, a population from which, in effect, the Janis-Feshbach variables were selected at random.

It is noted in passing that this reanalysis of the Janis-Feshbach results has followed the basic procedure used by the authors, that of comparing groups. An alternative procedure would have been to correlate talk-mediated anxiety (Variable 1 and/or 2) with each of the other six variables. It is too late to do this now, of course, since the decision to have proceeded thusly would have had to be made before the data were inspected. As it was, Janis and Feshbach chose, in effect, to assign talk-mediated anxiety scores based on the authors' pre-experimental evaluation of the fear-arousing effect of the talk. In effect, each student in a group (e.g., Strong) received the same score however much this experimenter-imposed score differed from the student's fear as indicated by his scores on Variables 1 and 2.

The outcome of this reanalysis of the Janis and Feshbach results suggests a caution in accepting the conclusions of Janis and Feshbach, which they purport to be supported by their results. For instance, on page 87 they state, "Thus, the findings consistently indicate that inclusion of the fear-arousing material not only failed to increase the effectiveness of the communication, but actually interfered with its over-all success." This is not to say that the defensive avoidance hypothesis or other conclusions of Janis and Feshbach are wrong - it's always possible to be right for the wrong reasons - but only to say that we must look to studies in this area which achieve significant results in order to assess the Janis-Feshbach hypotheses. It would seem more efficient to move on to these other studies than to cover the several pages of discussion by Janis and Feshbach (1953), for this discussion is based on their results, and their results, taken as a unit, are not significant. The only portion of their discussion which will be indicated here is that portion directly related to the primary interest of the current study - namely, the effect of arousal on learning and recall.

It would seem that Janis and Feshbach take the position that, under the conditions of their study, arousal has no effect on learning, but does on retention such that arousal and retention are negatively correlated.

First, the Janis-Feshbach position on learning. On page 88 they state, "Our results provide no evidence that a strong fear appeal produces inattentiveness or any form of distraction that would interfere with learning efficiency during the communication session." On page 89 they state, "Our findings definitely suggest that the use of fear-arousing material of the sort presented in the illustrated talks would rarely give rise to any interference with the audience's ability to learn the content of the communication." On page 92 the authors state, "The evidence indicates that the emotional reactions aroused by the Strong appeal did not produce inattentiveness or reduce learning efficiency."

And now, the Janis-Feshbach position on recall. On page 87 they indicate that students in the higher fear groups "...were inclined to avoid recalling [one week after the communication] the content of the fear-arousing communication." On page 90 they state, "It would be expected that those students [allegedly those in the lower fear groups] who changed their practices, after having heard and seen one of the three forms of the illustrated talk, were motivated to do so because they recalled some of the verbal material which had been given in support of the recommendations, most of which referred to unfavorable consequences of continuing to do the 'wrong' thing." On page 91 the authors indicate that when threatening material is learned under conditions of emotional tension, "... the audience will become motivated to avoid recalling those statements on later occasions when appropriate action could ordinarily be carried out." On page 91 the authors indicate that students in the higher fear groups revealed "...a tendency to avoid recalling the content of the fear-arousing communication."

For their measure of learning, Janis and Feshbach (1953) appear to be using immediate recall of information (Variable 3 in Table 1.7), and for their measure of recall, they appear to be using student responses to one open-ended question presented immediately after exposure to counter-propaganda (one week after the communication). The distinction they make between learning and recall will be considered here as a distinction between immediate and delayed recall. Using this latter distinction, the Janis and Feshbach position with respect to what they term "learning" and "recall" might be summarized as follows: arousal during exposure to a communication has no effect on immediate recall, but has a negative effect on delayed recall (higher arousal leads to poorer delayed recall).

The relationship stated in the last sentence might be correct, of course, but in view of the outcome of the preceding reanalysis of the

Janis-Feshbach results, it would seem appropriate to review other studies bearing on the relationship between arousal and recall. Two additional studies used the Janis-Feshbach treatments in modified form, as well as some of the Janis and Feshbach measures, and the following reviews will start with these two studies. The hypptheses of one study were in some cases different from those of the other, and also from those of Janis and Feshbach (1953). Of these hypotheses, the only ones which will be discussed below to any degree will be those pertaining to the primary interest of the current study - that is, the effect of arousal on retention.

1.13 Moltz and Thistlethwaite (1955)

This and the next study used the Janis-Feshbach treatments with three modifications: (1) the Moderate appeal was not used, (2) the contents of the talk and slides were modified, with nearly 50% fewer threats in the Strong and Minimal appeals than appeared in the original Janis-Feshbach versions, and (3) the material involving recommendations on proper dental hygiene practices received greater emphasis and was placed as a unit after the threatening material, whereas in the Janis-Feshbach versions, this material was interwoven with the threatening material.

Moltz and Thistlethwaite (1955) argued that the effectiveness of fear-arousing material is a function, in part, of the degree of anxiety reduction following the fear arousal. Thus, even the Strong fear appeal should be effective if it is followed immediately by material which assures the subject that he has available to him a course of action which can eliminate the basis for the fear. The authors used two measures of effectiveness: information retention and reported conformity.

In addition to a control group which will hereafter not be mentioned, Moltz and Thistlethwaite (1955) utilized Strong and Minimal groups, with each consisting of three subgroups. The first subgroup (No-Hygiene) received no material after the threatening material. The second subgroup (No-Assurance) received dental practices recommendations after the threatening material. The third subgroup (Assurance) also received the dental practices recommendations after the threatening material, but with explicit assurance as to the efficacy of the recommended practices in preventing tooth decay. The subjects were 329 newly-inducted Air Force recruits undergoing basic training.

Moltz and Thistlethwaite (1955) utilized only questionnaire measures, of which there were three: (a) a 14-item dental anxiety scale administered one week before, and immediately after, the communication, (b) a 12-item dental information scale administered one week before, and immediately after, the communication, and (c) a 4-item dental practices scale administered one week before, and one week after, the communication. This information is summarized in Table 1. 8.

TABLE 1. 8

VARIABLES UTILIZED IN THE MOLTZ AND THISTLETHWAITE STUDY (1955)

No.	Variable	No. of Items	1 Week Before	Just After	1 Week After
1	talk-unrelated dental anxiety	14	x	x	
2	talk-related dental information	12	x	x	
3	talk-recommended dental practices	4	x		x

The authors analyzed the dental anxiety scale for reproducibility, in the sense of Guttman (1950), and concluded (page 233) that "The results indicated extremely low levels of reproducibility when all 14 of the original items were considered."

The authors then compared the Strong and Minimal No-Hygiene subgroups on anxiety mediated by the two communications (the No-Assurance and Assurance subgroups could not be used in this comparison, since for these subgroups communication-mediated anxiety would be confounded with the communication recommendations or recommendations-assurance).

The authors found no significant difference between the No-Hygiene subgroups, and concluded (page 233) that "... the experimental treatments intended to produce anxiety arousal did not have the intended effect."

This is essentially the same result as was obtained by Janis and Feshbach. However, it does not follow necessarily that the experimental groups will fail to differ significantly in information recall, as was the case in the Janis-Feshbach study, for Moltz and Thistlethwaite modified the illustrated lectures so that the recommendations or recommendations-assurance followed the threatening material. In fact, it was the authors hypothesis that, following Janis and Feshbach (1953), "... greater anxiety reduction would be associated with significantly better learning and more conformity to recommendations contained in a communication." (Moltz and Thistlethwaite, 1955, page 231). Nevertheless, even though there was a tendency for the two Assurance subgroups, relative to the two No-Assurance subgroups, to show a lower anxiety immediately after the communication, presumably reflecting a reduction in anxiety as a result of the assuring material, these subgroups failed to differ significantly with respect to information retention.

Moltz and Thistlethwaite (1955) also hypothesized, following Janis and Feshbach (1953), that those subjects who received the material recommending certain dental practices (the No-Assurance subgroups) would show one week after the communication greater reported conformity, relative to the No-Hygiene subgroups, to the recommended practices. However, this hypothesis was not supported. The authors also hypothesized, following Janis and Feshbach (1953), that the Assurance subgroups would show greater conformity than would the No-Assurance subgroups, but this hypothesis, too, was not supported. The authors summarized (page 236) these results as follows: "Greater anxiety reduction was not associated with greater learning nor was it associated with greater reported conformity to the recommendations. Thus, neither prediction made on the basis of the present anxiety-reduction hypothesis was confirmed."

While this conclusion seems to be appropriate to the results of this study, one cannot be entirely certain because of an error made by Moltz and Thistlethwaite (1955) in some of their analyses. When three groups were involved in an analysis, these authors followed the Janis and Feshbach (1953) procedure of determining the null probability for each pair of groups. The three resulting probabilities were then combined by Moltz and Thistlethwaite by Equation 1. 4 of this report. The inappropriateness of this procedure has already been discussed in this report.

1. 14 Goldstein (1959)

The Goldstein study (1959) represents the third attempt to induce a differential arousal effect by means of the illustrated talk on dental hygiene. Goldstein utilized that version of the communication which was developed by Moltz and Thistlethwaite (1955).

Goldstein argued that the effectiveness of fear-arousing material is a function not only of the level of arousal stimulated by the appeal but also of the subject's characteristic reaction to tension-producing stimuli. Thus, if the subject tends to recall minimally-arousing material better than strongly-arousing material, he should respond more favorable to the Minimal appeal. If, on the other hand, the subject tends to recall strongly-arousing material better than minimally-arousing material, he should respond more favorably to the Strong appeal. The subject's characteristic mode of handling tension-producing stimuli was measured by a variation of the Sentence Completion Test consisting of sentence stems involving sexual and aggressive implications; subjects who selected completion choices in terms of their own needs and emotions were classified as Copers, whereas subjects who selected the less personal completion choices were classified as Avoiders. The study involved 67 Copers and 72 Avoiders, all high school freshmen.

The condition of the study were similar to that of Janis and Feshbach (1953), with the exception that the treatments differed considerably from those of Janis and Feshbach, as noted in the preceding review of the Moltz and Thistlethwaite study (1955). In view of these differences, it is surprising to find the statement by Goldstein (1959) that his materials "... represent a slight modification of the original appeals used by Janis and Feshbach."

Further, while Goldstein states that the two "... appeals contained the same information about the causes of tooth decay and they both contained identical recommendations," a review of the lectures (Goldstein, 1957) reveals vast differences in these regards. For instance, the Strong appeal contained 7 paragraphs and 5 slides consisting of information and recommendations pertaining to proper tooth-brushing practices, but this material was completely lacking in Goldstein's Minimal appeal. The Strong appeal consisted of about 2400 words, the Minimal appeal about 1500.

One might suspect that the basis for the above discrepancy arises from a failure here to recognize that Goldstein (1959) is referring to a Janis and Feshbach study different from the one discussed at length in this report. Such a suspicion could arise, for Goldstein (1959) related his study to a previous study which he cites seven times as being reported in Feshbach and Janis (1954) and Janis and Feshbach (1955). However, both references are incorrect, and it appears fairly clear that Goldstein (1959) really had in mind the Janis-Feshbach study discussed in this report, and referenced here as Janis and Feshbach (1953), and Janis and Feshbach (1954).

Table 1. 9 shows the three Goldstein variable which are of interest here.

TABLE 1. 9

VARIABLES UTILIZED IN THE GOLDSTEIN STUDY (1959)

No.	Variable	No. of Items	1 Week Before	Just After	2 Weeks After
1	talk-unrelated dental anxiety	16*	x	x	
2	talk-related dental information	22	x	x	x
3	talk-recommended dental practices	5	x		x

*reduced to 15 items before analysis

In order to determine whether the two treatments induced a differential experimental effect, Goldstein administered a 16-item dental anxiety scale, with one of the items deleted subsequently because the students had difficulty in handling it. This dental anxiety scale was administered immediately after the lecture as well as sometime before the lecture. The interval between the first administration and the lecture is stated by Goldstein (1959) as one week, but a subsequent article (Goldstein, 1960) gives this interval as two weeks; the shorter interval is shown in Table 1. 9.

The reader gets the impression that when Goldstein (1959) refers to the Strong fear appeal (21 references, and the Minimal fear appeal (15 references), he is referring to two appeals which differ in the amount of fear they induce. Goldstein, following Janis and Feshbach (1953), gives the strong impression of using the terms "Strong" and "Minimal" in a quantitative sense and with reference to fear (or tension, anxiety, arousal), an impression strengthened by Goldstein's use of such phrases as "level of fear arousal," "level of tension," "level of emotional tension" (7 references).

Further, although Goldstein (1959) does not state so explicitly, he appears to imply that he achieved a significant treatment effect, with the Strong group showing greater communication-mediated arousal. For instance, Goldstein (1959, page 247) states that the effectiveness of a communication is a function of the "... interaction between the person's characteristic mode of responding to tension-arousal and the level of arousal stimulated by the appeal." On page 248 Goldstein predicts "... an interaction between personality type and level of fear arousal." On page 251 the author states that "The two levels of fear-arousal do stimulate differential acceptance among copers and avoiders," On page 251, Goldstein states that "Although an interaction between a particular personality variable and the level of emotional tension stimulated has been found in the present study..."

However, Goldstein (1959) presents no results bearing directly on the question of experimental effect, forcing the reader to the more complete account of the study (Goldstein, 1957). The experimental effect is indicated on page 26, which shows a lecture effect F of .13, with pre-communication dental anxiety as a covariate, and the following sentence: "An incidental finding, is the absence of significant difference on dental anxiety scores between the subjects who received the two appeals." In view of the fact that Goldstein's hypothesis pertains to the interaction between characteristic reaction and arousal level with respect to conformity, the absence of an arousal effect is about as incidental as, say, the use of females only in the test of an hypothesis pertaining to an interaction between grade level and sex with respect to IQ.

Janis and Feshbach (1953) failed to show an experimental effect, and on this basis it was predicted in this report that these authors would fail to demonstrate an other-than-chance over-all difference among their three experimental groups with respect to other measures. It has been shown in this report that this prediction was correct. Next, Moltz and Thistlethwaite (1955) failed to show an experimental effect, and they too failed to find a significant difference between their two experimental groups with respect to their two effectiveness measures: retention and conformity. Since Goldstein (1957) also failed to induce an experimental effect, one would again be surprised if other-than-chance differences emerged between his two experimental groups with respect to either effectiveness measure.

Following Moltz and Thistelthwaite (1955), Goldstein utilized two types of effectiveness measures: retention (immediate and delayed) and conformity. The two experimental groups failed to differ significantly with respect to either immediate or delayed retention, as measured by recognition - not recall, as indicated 11 times by Goldstein (1959) - and controlled for pre-communication dental information (Goldstein, 1957).

The two experimental groups also failed to differ significantly with respect to conformity, as measured by before-after change in dental practices. Goldstein employed the three types of conformity measures which were previously utilized and defined by Janis and Feshbach (1953): (1) increased conformity, (2) decreased conformity, and (3) no change. The frequencies in these three categories are given in Table 1.10, which also shows the net effects in the last row. Goldstein does not compare the net effects given in Table 1.10, but elsewhere he compares net effects by utilizing the equations given in Hovland, Lumsdaine and Sheffield (1949, page 304), Following the same procedure with the data shown in Table 1.10, one obtains a t-ratio of 1.65, which is not significant for the difference between the Strong and Minimal groups with respect to conformity net effect.

TABLE 1.10

FREQUENCIES OF CHANGE IN REPORTED DENTAL PRACTICES
FOR EACH EXPERIMENTAL GROUP IN THE
GOLDSTEIN STUDY (1959)

Type of Change	Strong	Minimal	Total
Increased Conformity (A)	29	31	60
Decreased Conformity (D)	17	8	25
No Change (B+C)	30	24	54
(A-D)	12	23	35
(a-d)	.16	.37	.25

Inasmuch as Goldstein (1959) failed to demonstrate an experimental effect (a significant difference between the Strong and Minimal groups with respect to arousal), it comes as no surprise to find that these two groups also failed to differ significantly with respect to retention (immediate or delayed) and conformity. However, it does come as a surprise to find Goldstein reporting a significant interaction between characteristic reaction and arousal level with respect to conformity. As indicated previously, this is analogous to reporting a significant interaction between grade level and sex with respect to IQ on the basis of a study involving females only.

Unfortunately, Goldstein (1959), who basis his report of a significant interaction on an unpublished statistical test of significance, fails to supply to the reader any particulars about the test. Goldstein (1959, page 250) simply states that "The critical ratio of this second order effect is 1.60, which is significant to the .05 point for a one-tailed test of significance," and adds to this sentence the following footnote: "Special thanks are due Fred Sheffield for informing the author of the method of extending the test of significance for net effect to the second order comparisons." A first-order comparison pertains to the difference between the Strong and Minimal groups with respect to net effect; a second-order comparison pertains to the difference between two groups (Copers and Avoiders) with respect to the difference between the Strong and Minimal subgroups with respect to net effect. Thus, some of the Copers were exposed to the Strong appeal, while the remaining Copers were exposed to the Minimal appeal, and these two subgroups will show a difference with respect to net effect; a similar situation holds for the Avoiders. Goldstein's hypothesis predicts that, relative to the difference between the two Coper subgroups with respect to net effect, the difference between the two Avoider subgroups with respect to net effect will be significant. Since Goldstein (1957, 1959) has not presented the details of the analysis pertaining to his hypothesis, it is appropriate to do so here.

The second-order effect is computed from the subgroup frequencies, which are given in Table 1.11. As usual, the net effect is given in the next-to-the-last row in terms of frequencies, and in the last row in terms of proportions. Goldstein's test of interaction pertains to the question of whether, relative to the difference between the Coper subgroups with respect to net effect (.281-.257), the difference between the Avoider subgroups with respect to net effect (.452-.073) will be significant. The standard error of the difference between these differences was obtained by Goldstein (1959) by extending the equations given in Hovland, Lumsdaine, and Sheffield (1949), and reproduced here as Equations 1.2 and 1.3. Under the null hypothesis that the four subgroups are from the same population, the combined frequencies indicated in Equation 1.3 are shown in the right-most column

TABLE 1. 11

FREQUENCIES OF CHANGE IN REPORTED DENTAL PRACTICES FOR EACH EXPERIMENTAL SUBGROUP IN THE GOLDSTEIN STUDY (1959)

	Copers		Avoiders		
Type of Change	Strong	Minimal	Strong	Minimal	Total
Increased Conformity (A)	15	14	14	17	60
Decreased Conformity (D)	6	5	11	3	25
No Change (B+C)	14	13	16	11	54
(A-D)	9	9	3	14	35
(a-d)	.257	.281	.073	.452	.252

of Table 1. 11. Equation 1. 2 gives the first-order error variance (for the difference between the two Coper subgroups, or the difference between the two Avoider subgroups); the second-order error variance is obtained by extending Equation 1. 2 as the sum of four terms, which is the same as utilizing Equation 1. 3 with the reciprocal frequency summed over the four subgroups. The value of this second-order standard error is .253. Utilizing this value as the error term,

$$t = \frac{(.452-.073) - (.281-.257)}{.253} = 1.40,$$

which is not significant.

The obvious question which now arises pertains to the discrepancy between the report of a significant interaction by Goldstein (1959) and the report here of a non-significant interaction. Goldstein bases his significant interaction on a t-ratio of 1. 60, and in the absence of pertinent information in the article (Goldstein, 1959) or in the full report (Goldstein, 1957), one cannot be entirely certain where Goldstein made his mistake. However, it is interesting to note that if one reverses the second difference in the numerator of the equation above, so that it becomes (.257-.281), a t-ratio of 1. 60 is obtained. Without necessarily assuming that Goldstein made this particular error, the inappropriateness of such a reversal should be apparent.

In order to test for interaction, the difference between the Avoider subgroups must be compared against the same difference between the Coper subgroups, not between the reversed difference between the Coper subgroups. If, incorrectly, the latter procedure were applied, the interaction would appear to increase as the difference between the Coper subgroups increased and became equal in magnitude and direction to the difference between the Avoider subgroups.

One might suspect that the discrepancy between the t-ratio as computed by Goldstein (1959) and as recomputed here might be due to a different organization by Goldstein of the frequencies given in Table 1. 11. One could, of course, consider the main groups to be Strong and Minimal, and the subgroups to be Copers and Avoiders. However, such a reorganization of Table 1. 11 would not affect the interaction.

Another possible basis for the discrepancy might be the use by Goldstein of a second-order error variance based on the sum of two first-order error variances. This is not recommended for two reasons. First, the combined proportions implied in Equation 1. 3 would be based on less than the total sample frequencies available, and, therefore, would tend to be poorer estimates of the population proportions. Second, a unique value of the second-order error variance could not be obtained, because the values of the first-order variances would depend somewhat on whether Table 1. 11 were organized as shown, or with Strong and Minimal considered as the main groups. If organized as shown, a t-ratio of 1. 40 is obtained for the interaction; if organized the other way, the t-ratio becomes 1. 43. Thus, it is clear that Goldstein did not obtain his t-ratio of 1. 60 by either of these alternate methods.

The hypothesis that Goldstein made his error by reversing the difference between the Coper subgroups gains support by the fact that Goldstein (1959) converts his incorrect t-ratio to a probability level on the basis of one tail of the normal distribution, and, incidentally, reporting incorrectly and in the direction of his hypothesis the probability level associated with a t-ratio of 1. 60 with 135 df. Only by reversing the net effects (last row of Table 1. 11) of the two Coper subgroups would Goldstein, according to his own argument, be justified in using a one-tailed test. Goldstein (1957, page 24) states that "A one tailed test was used because the nature and the respective positions of each group were predicted in advance."

This commonly-employed rationale for the use of the one-tailed test is incorrect. A one-tailed test is appropriate when the investigator is interested only in an unidirectional effect. Prediction of direction of results is not an appropriate basis for the utilization of a one-tailed test. In fact, to base a one-tailed test on the assumption that the alternative result cannot emerge represents a presumptuousness which is likely to be deflated in time by nature, which seems to abhor the assumption of complete precognition.

However, even if Goldstein were correct in assuming that prediction of direction allows the application of a one-tailed test, it would not be sufficient. An additional requirement would be that the results emerge in the predicted direction, and Goldstein's results failed to meet this second requirement. On page 247 Goldstein (1959) gives his interaction hypothesis, as well as the two predictions arising from the hypothesis.

> "It is hypothesized that the acceptance or non-acceptance of the recommendations contained in a propaganda appeal is related to the S's characteristic reaction to tension-producing stimuli. If an S tends to recall neutral stimuli better than tension-producing stimuli, he should respond more favorably to a minimal fear appeal, in which the recommendations are introduced in a relatively neutral context. If, on the other hand, he tends to recall tension-producing stimuli better than neutral stimuli, he should respond more favorably to a strong fear appeal, in which the recommendations are introduced in of setting a heightened emotional tension."

The last sentence, of course, pertains to Copers, and it is clear from the last row of Table 1.11 that Copers responded less favorably to the strong fear appeal. Thus, the reversal of the net effects for the two Coper subgroups would not only bring Goldstein's data in line with his predictions but also justify his use of a one-tail test, at least according to his incorrect rationale regarding one-tailed tests.

It is now possible to summarize the one"significant" result reported by Goldstein (1959). The interaction between characteristic reaction and arousal level with respect to conformity can be considered significant, as purported by Goldstein (1959), provided that the reader is willing to accept Goldstein's computational error and his utilization of a one-tailed test using data which contradict predictions in part and his liberal interpretation of the probability level associated with the incorrect t-ratio In addition, the reader must be willing to recognize the fact that since Goldstein (1959) reported a vanishingly-small difference in arousal level, only a chance interaction would be expected.

Despite the absence of a single valid significant result, Goldstein (1959), pages 251 and 252) claims that his study

> "... leaves the conclusions of Janis and Feshbach largely unchanged. A minimal fear appeal is still most effective in eliciting acceptance of propaganda. It is not particularly effective with persons classed as copers yet it does not alienate, as does the strong fear appeal, the avoiders. There does not appear to be any great advantage in tailoring the level of fear stimulation contained in a propaganda appeal to different personality types. Unless some form of propaganda can be found that is particularly effective with the coper groups, a minimal fear appeal still stands as the best bet for the propagandist."

This conclusion of Goldstein (1959) has been criticized by Janis and Terwilliger (1962, page 403), who note that with regard to communication effectiveness, Goldstein "... reports a difference between the strong and mild fear appeal groups (without presenting the significance test data) and claims that it supports Janis and Feshbach's conclusion concerning the greater effectiveness of a minimal appeal." However, Janis and Terwilliger (1962), realizing that Goldstein failed to obtain a significant differential arousal between his two groups, conclude therefore that the Goldstein study cannot "...be regarded as providing an adequate test of any hypothesis concerning differences in amount of attitude change induced by different degrees of fear or anxiety arousal."

Incidentally, the absence of the significance test noted by Janis and Terwilliger (1962) has been supplied in this report. It is recalled that the result indicated that the Strong and Minimal groups do not differ significantly with respect to communication effectiveness.

Since the primary interest in the current study pertains to the effect of arousal on information retention, it is appropriate to review in some detail Goldstein's discussion of the influence of arousal level on information retention.

Goldstein (1957, page 8) predicted that type of appeal (Strong or Minimal) would have no influence on information retention immediately after the presentation. Since a prediction phrased in this manner, even if confirmed by the data, lends itself to so many interpretations, Goldstein's failure to find a significant difference between the appeals with respect to immediate retention cannot be used as critical evidence to support his theoretical basis for such a prediction.

Goldstein (1959) predicted that a Coper would show greater acceptance of a message which induces a higher level of arousal because of a greater retention of such a message, relative to a low-arousal message. Similarly, an Avoider would show greater acceptance of a message which induces a low level of arousal because of a greater retention of such a message, relative to a high-arousal message. Thus, retention is conceptualized as the mechanism for acceptance. In this regard, Goldstein (1959) follows Janis and Feshbach (1953), who postulate that the failure of a threatening message to have a delayed effect is due to the inability of the communicator to recall the content of the communication at a later time. In order to test this postulation, Goldstein (1959) obtained a delayed measure of information retention, a measure not obtained by Janis and Feshbach (1953) or Moltz and Thistlethwaite (1955).

The differential prediction with regard to delayed retention was not supported by the data, and Goldstein suggests that this failure may be due to the fact that he collected the retention data in the classroom rather than in the bathroom. He conjectures that, unlike Copers, Avoiders may be unable to recall the lecture information at the time of teeth brushing, whereas both personality types may be capable of showing a retention of the information in the classroom. Thus, Goldstein (1957, page 40) states that "Possibly, both copers and avoiders recall the content of the fear arousing lecture when in the classroom situation. However, only the coper group may be capable of recalling the content of the lecture when about to brush their teeth. Such a postulation would account for Goldstein's failure to obtain a retention difference in the classroom, but in absence of supplementary bathroom data, Goldstein's classroom-bathroom hypothesis has yet to be fully verified. In short, Goldstein's hypothesis is compatible with, but is not required by, his data, which shows only that the two treatment groups, as well as the two personality types, failed to differ significantly with respect to retention in the classroom.

Goldstein (1959) argued that if retention is the mechanism for conformity, retention and conformity should be positively related. The correlation coefficient would have constituted the simplest and most direct statistic to test this hypothesis. Instead, Goldstein tested the difference between Copers and Avoiders with respect to information retention. This procedure is less direct and less sensitive, and further requires as a minimum, the fulfillment of each of three conditions: (1) Copers and Avoiders must differ significantly with respect to retention, (2) Copers and Avoiders must differ significantly with respect to conformity, and

(3) the group (Copers or Avoiders) which is higher on retention must be the same group which is higher on conformity. Oddly enough, Goldstein (1957, 1959) made no test of (2); it is not clear whether he assumed that one of the two groups was significantly higher than the other, and, if so, which of the two groups. Restricting his analysis to (1), and finding no significant difference between Copers and Avoiders with respect to retention, Goldstein (1959, page 250) concludes: "Thus, the hypothesis that differences in acceptance behavior [conformity] are mediated by differential recall of the content of the appeals is not substantiated." Goldstein (1959, page 251) discusses the implications of his conclusion.

> "Although the prediction of differential acceptance behavior by copers and avoiders was based on assumption of differential recall patterns, the analyses of learning and retention of content provide no evidence of differential recall. The mediating mechanism for the responses of copers and avoiders thus remains obscure. This finding is particularly interesting in view of the weight originally placed by Janis and Feshbach (1955) upon the role of recall in explaining the failure of the strong fear appeal. If there is no relationship between the recall of the content of a propaganda appeal and its acceptance, then it becomes necessary to search for other explanations for the basis of acceptance behavior."

However, since the validity of Goldstein's argument requires not simply the prediction of differential conformity by Copers and Avoiders but also the demonstration that Copers and Avoiders do in fact differ significantly with respect to conformity, and since Goldstein (1957, 1959) fails to supply to the reader the results of such a test, that test will have to be made here. Inasmuch as the Strong and Minimal groups failed to differ significantly on any measure, these two groups will be combined so as to allow a simple comparison of Copers and Avoiders.

Table 1.11 indicates that 29 Copers and 31 Avoiders showed increased conformity, while 11 Copers and 14 Avoiders showed decreased conformity. Applying the net effect analysis used by Goldstein, the difference between Copers and Avoiders with respect to conformity is significant at the .80 level. Thus, Goldstein's argument that retention is not the mediating mechanism of communication effectiveness appears to be without foundation.

In view of the fact that Copers and Avoiders fail to differ significantly with respect to retention and conformity, one suspects also that they might not differ significantly with respect to anxiety. Goldstein (1957, page 26) shows that, in fact, this is the case. Further, there were no significant results for the personality-by-lecture interaction with respect to (1) dental anxiety, (2) immediate retention, and (3) delayed retention (Goldstein, 1957, pages 26-28).

The Goldstein study is readily summarized. In a 2x2 design Copers and Avoiders were exposed to Strong and Minimal fear communications. Four measures were obtained from each student: anxiety, immediate retention, delayed retention, and reported conformity. The results revealed that Copers and Avoiders failed to differ significantly with respect to any measure, that the Strong and Minimal groups failed to differ significantly with respect to any measure, and that there were no significant interactions with respect to any measure.

On the basis of these results Goldstein (1959) concluded that "The superiority of the minimal fear appeal with avoiders is clearly demonstrated..." (page 251); "...an interaction between a particular personality variable and the level of emotional tension stimulated has been found in the present study..." (page 251); "Unless some form of propaganda can be found that is particularly effective with the coper groups, a minimal fear appeal still stands as the best bet for the propagandist" (page 252); "The results support the hypothesis that a strong fear appeal receives greater acceptance among copers than among avoiders [p=.31], while the minimal fear appeal receives greater acceptance among avoiders than among copers [p=. 34]" (page 252); and "... the concept of a 'defensive reaction' to the arousal of strong emotional tension, introduced by Janis and Feshbach, is limited in its explanatory power" (page 252). Elsewhere Goldstein (1957) concluded on the basis of his results that "if the strong appeal can be considered to be a stressful situation, avoiders show a greater disruption of behavior than do copers under such conditions" (page 38); "Despite the differential response of copers and avoiders to the two appeals, the results indicate that there is little value in using propaganda which attempts to stimulate a high degree of emotional tension" (page 39); "As far as the results of this study can be generalized to other populations and other topics, a minimal fear appeal is most effective (page 39); "Possibly, avoiders can recall the content of the lectures in any situation but the one in which they are about to brush their teeth" (page 40); and "The present study confirms the hypothesis that the reaction of a subject to fear-arousing propaganda is related to his characteristic method of handling anxiety" (page 46).

1. 15 Berkowitz and Cottingham (1960)

Following a pilot study, which will not be reported here, Berkowitz and Cottingham (1960) performed a study to test, allegedly, the following two predictions: (1) an uninteresting Minimal fear appeal will not mediate an attitude change, and (2) an interesting Strong fear appeal will mediate an attitude change which is inversely proportional to the relevance of the message for the subject.

The subjects were 64 volunteers from introductory psychology college courses. In groups of at least 18, they listened to a tape-recorded lecture which advocated the utilization of automobile safety belts. The study involved two lecture conditions, Strong and Minimal fear arousal. The Strong lecture was heard by 26 students, the Minimal lecture by 20 students, with the remaining 18 students acting as controls.

Six measures were obtained from the experimental subjects immediately after the lecture. (1) Anxiety was measured by the question, "I would classify the presentation as (a) very pleasant, (b) pleasant, (c) nonstimulative, (d) unpleasant, or (e) very unpleasant." (2) Appraisal of interest value was measured by the question,"The presentation of the material was (a) boring, (b) of indifferent quality, (c) thought provoking, or (d) disturbing." (3) Aggression toward the communicator was measured by one item which asked the student to indicate whether the lecturer sounded (a) very pleasant, (b) pleasant, (c) neutral, (d) unpleasant, or (e) very unpleasant. (4) Information retention was measured by an 11-item information scale. (5) Relevance of the communication for the student was measured by car ownership and usage: the message had high relevance for a student who owned a car and drove it at least several times a week, medium relevance for a student who owned a car but drove it less frequently, or did not own a car but rode in one at least several times a week, and low relevance for a student who neither owned a car nor rode in one more than once a week or on weekends. (6) Attitude change toward safety belts, the effectiveness, criterion, or dependent variable, was measured by the student's before-after change in response to a scale consisting of 18 agree-disagree statements, with the pre-communication response being obtained one month before the treatment. Only the last two measures were obtained from the control subjects; the first four were not appropriate since they pertained to the communication. These six variables are summarized in Table 1.12.

Berkowitz and Cottingham (1960) imply that they are using the terms Strong and Minimal anxiety arousal in a quantitative sense, as was true in the three previous studies; presumably, then, the two groups should differ significantly with respect to anxiety (Variable 1). Unfortunately, there is no way of determining the difference between these two groups, because Berkowitz and Cottingham (1960) have chosen to edit the results which are presented to the reader. The anxiety item consists of five ordered

TABLE 1.12

VARIABLES UTILIZED IN THE BERKOWITZ AND COTTINGHAM STUDY (1960)

No.	Variable	No. of Items	1 Month Before	Just After
1	anxiety induced by communication	1		x
2	appraisal of interest value	1		x
3	aggression toward communicator	1		x
4	information retention	11		x
5	relevance of communication	2		x
6	attitude change (toward safety belts)	18	x	x

categories, but Berkowitz and Cottingham (1960) show only the three middle categories, with all 46 experimental subjects entered in these three categories. The authors do not indicate whether the two extreme categories are not shown because they contained no entries or because their entries were combined with adjacent categories. In any case, the frequencies shown in the three categories reveal no significant ordered difference between the Strong and Minimal groups with respect to anxiety: by combining two of the categories in order to eliminate the extreme non-ordering of frequencies which exists within each group, since the categories are purported to be ordered, the resulting 2x2 table leads to a chi square of 3.18, which is not significant. Thus, it appears that the Berkowitz and Cottingham study becomes still another study reviewed here which fails to demonstrate an experimental effect; apparently this phenomenon is not simply a reflection of the use of the tape-slide presentation on dental hygiene.

Berkowitz and Cottingham (1960) also follow an unusual procedure with regard to their presentation of responses to the item intended to measure interest value (Variable 2). Of the four categories for this item, the authors do not publish the responses in one of the categories, and further, combine two of the remaining categories. No reasons for these procedures are offered, and the reader cannot really determine whether the Strong and Minimal groups differ significantly with respect to interest value. Such a determination is not particularly important anyway, for the authors decided to introduce this variable into their study after they had analyzed the responses to this one item. Before having seen these responses, the authors had intended to consider this as a second item in the measurement of anxiety. Berkowitz and Cottingham (1960) describe their reasons for deciding to introduce an interest value variable in their study (Experiment II) even though it had not been anticipated in advance on basis of the results of the pilot study (Experiment I).

> The responses to the items were combined in Experiment I by weighting each alternative in terms of the amount of discomfort indicated and adding the two scores. The higher the score, then, the higher the admitted tension evoked by the communication. However, in Experiment II the items were kept separate for two reasons. First, answers to the items exhibited a somewhat different pattern in the second study. Fewer people in the latter investigation described themselves as very upset by the lecture, particularly in response to the second question, while a greater proportion of the Ss indicated the material was uninteresting to them. Second, the greater frequency of responses indicative of boredom in Experiment II made it possible to test the effects of communication interest-value upon opinion change.

Since Berkowitz and Cottingham readily admit that they had not intended to measure interest value, it is obvious that they could not have predicted that effectiveness would be positively related to interest value. Since Berkowitz and Cottingham agree with Janis and Feshbach that high arousal reduces communication effectiveness, it must be assumed that Berkowitz and Cottingham actually predicted that the Minimal appeal would be more effective.

But even if Berkowitz and Cottingham had in fact decided to utilize an interest variable before looking at their results, it is not at all clear why these authors would predict a greater effectiveness for the Strong appeal simply because it is more interesting. The basis for such a prediction is not clear because Berkowitz and Cottingham claim to adhere to the thinking of Janis and Feshbach, and Janis and Feshbach state that a Strong appeal can be expected to be more interesting and less effective (Janis and Feshbach, 1953, page 82). Thus, the Berkowitz and Cottingham finding of a greater interest for the Strong appeal simply corroborates the Janis and Feshbach conclusion.

Berkowitz and Cottingham (1960) present no quantitative data regarding differences between the two experimental groups with respect to aggression toward communicator (Variable 3) or retention (Variable 4), but indicate simply that the two groups did not differ significantly with respect to these two variables.

The apparent failure of the two experimental groups to differ significantly with respect to the first three variables would probably not come as a surprise to the reader who had been wondering why an item which asks whether the presentation was pleasant constitutes a measure of anxiety arousal, whereas an item which asks whether the taped lecturer sounded pleasant constitutes a measure of aggression toward the lecturer. Such an unimaginative reader might also have wondered whether it wouldn't have been more direct to measure anxiety by the item for which "disturbing" constitutes the positive choice, and interest value by the item for which "very pleasant" constitutes the positive choice, rather than the reverse procedure, as followed by Berkowitz and Cottingham (1960). The naive reader might even wonder why anxiety was not measured by an item which simply asks the student to indicate his degree of anxiety on a 5-point scale, and why interest was not measured by an item which simply asks the student to indicate his degree of interest on a 5-point scale. Except when it is clearly inappropriate, the direct method is likely to result in a more reliable and valid measure. The investigator who arbitrarily uses the indirect method may find himself in the position of the clinician who, according to a story by Allport (1953), was told by his patient that a particular Rorschach card made him think of sex. The clinician, thinking to tap a buried complex, asked him why. "Because," said the patient, "I think of sex all the time."

We turn now to the last two variables, relevance and attitude change, and to the two purported predictions made by Berkowitz and Cottingham (1960). The test of each prediction requires one and only one probability value, neither of which is supplied by the authors. Nevertheless, it appears that both predictions are fulfilled: (1) the Minimal and Control groups do not differ significantly with respect to attitude change, and (2) the three levels of relevance within the Strong group do differ significantly with respect to attitude change. Such are the results: the interpretations are less clear.

The first prediction is essentially meaningless; there are simply too many reasons why two groups might fail to differ significantly to justify reliance on the one reason offered by the investigators - in this case, lack of interest mediated by the Minimal appeal.

The second prediction, according to the authors, is a reflection of the defensive avoidance hypothesis. Following Janis and Feshbach (1953), Berkowitz and Cottingham (1960) assume that anxiety reduces the effectiveness of a communication. They also assume that higher relevance results in higher anxiety. Thus, they reason that higher relevance should reduce communication effectiveness. If it were true that higher relevance leads to higher anxiety, then relevance and anxiety would be positively and significantly related, which, according the authors (page 42), they are not. The absence of any simple relation is also suggested, but not critically demonstrated, by the frequency data of Berkowitz and Cottingham (1960). The authors classify the 26 students in the Strong group first according to three levels of anxiety (Variable 1) then according to three levels of relevance (Variable 5). The resulting frequencies, shown in their Tables 3 and 5, are reproduced here as Table 1. 13. If anxiety and relevance were positively related, one might have expected the two columns of frequencies to show more similar distributions than they do.

TABLE 1.13

FREQUENCIES FOR THREE LEVELS OF TWO VARIABLES FOR THE STRONG GROUP

	Anxiety	Relevance
High	8	11
Medium	2	9
Low	16	6

In sumary of this point, within the Strong group there is a negative relation between relevance and attitude change, but the basis for this relation remains unclear. However, the existence of this relation does not appear to support the defensive avoidance hypothesis (i. e., anxiety reduces communication effectiveness) because (1) Berkowitz and Cottingham state that the relation between anxiety and relevance was not significant for students within the Strong group, and (2) the authors fail to present any data which indicate a significant negative relation between anxiety and attitude change.

Finally, and rather surprisingly, Berkowitz and Cottingham (1960) completely fail to mention that the Strong group was significantly higher than the Minimal group with respect to attitude change in the communication-recommended direction.

1.16 Janis and Terwilliger (1962)

Janis and Terwilliger (1962) conducted a study designed to allow a measure of arousal at discrete points during the presentation of a communication. They asked 31 college and middle-age subjects individually to read either a High Threat or Low Threat communication concerning smoking and lung cancer. Each subject was asked to verbalize his reactions to the material at the end of each paragraph, as well as at any other time if he so chose. The subject could not hear his own verbalization because of noise presented to him through earphones.

A content analysis of the recorded verbalizations revealed that the High Threat subjects were associated with a significantly higher mean number of paragraphs that evoked expressions of affective arousal. Thus, of those studies reviewed so far, this becomes the first to demonstrate clearly an experimental effect. The two versions of the communication differed only in that the High Threat communication contained 7 additional paragraphs on the seriousness of lung cancer. The difference between the two groups with respect to affective reactions was primarily with respect to these 7 paragraphs. It is interesting to note that the first study reviewed here which achieved an experimental effect is also the first study to obtain measures of anxiety during the communication, not simply before and/or after.

Immediately before being presented the communication, the subject was asked to verbalize his thoughts as he imagined himself smoking. Then after the communication and a brief rest (the length of this rest is not given), the subject was again asked to verbalize his thoughts as he imagined himself smoking. The difference between these two verbalizations was taken as a measure of attitude change toward smoking. The authors apply an appropriate test of the difference between the groups with respect to attitude change. This difference is significant at the .21 level, a level slightly lower than that given by the authors; they apparently neglected to take into account the small size of their groups.

The primary purpose of the Janis and Terwilliger study was to test the resistance hypothesis that, relative to the Low Threat group, the High Threat group will display more resistance, which will be manifested in higher critical reactions during exposure to the communication. Critical reactions were measured by the mean (with respect to the number of subjects in each group) number of unfavorable statements made (or favorable statements not made) during the communication with regard to the 15 paragraphs common to both versions (the 7 supplementary paragraphs included in the High Threat version were not involved in this analysis). A count was made of evaluative comments in four categories: (1) unfavorable comments about content, (2) unfavorable comments about presentation, (3) favorable comments about content, and (4) favorable comments about presentation.

The authors compared the two groups with respect to each category separately. The High Threat group gave a higher mean number of evaluative comments to categories (1) and (3), whereas the Low Threat group gave a higher mean number of evaluative comments to categories (2) and (4). The over-all result is somewhat difficult to interpret because two categories are labelled unfavorable and two favorable. The t-ratio for categories (3) and (4) computed by the authors would remain unchanged if the mean number of favorable comments were replaced by its complement with respect to 15. This complement may be thought of as a relative unfavorable score, representing as it does the number of paragraphs for which favorable comments were not made. By complementing (3) and (4) thusly, and summing over the four categories, the total mean number of unfavorable paragraphs is obtained. With regard to this measure, the difference between the groups turns out to be 1.57. (Of course, if we had complemented categories (1) and (2) instead, the difference between the groups would still have been 1.57). The High Threat group gave a higher mean number of unfavorable comments, and this difference between the groups appears to be near the .05 level of significance (a test cannot be made on the basis of the information available). The authors point out that the difference between the groups was a function primarily of comments of the first type: unfavorable comments about content.

In further investigation of the resistance hypothesis, Janis and Terwilliger (1962) also classified comments regarding the 15 common paragraphs in a category called "Paraphrasing of arguments," statements made by the subject which paraphrase the recommendations presented by the communication. Such statements were considered by the authors as indicative of acceptance of the communication's recommendations. As expected by the authors, the High Threat group gave such statements to a lower mean number of paragraphs, and the difference between the groups approached the .05 level of significance.

The variables utilized in this study are indicated in Table 1.14.

TABLE 1.14

VARIABLES UTILIZED IN THE JANIS AND TERWILLIGER STUDY (1962)

No.	Variable	Just Before	During	Just After
1	affective arousal		x	
2	attitude change (toward smoking)	x		x
3	criticism of communication		x	
4	paraphrasing communication recommendations		x	

In summary of this study, it appears that when a standard message on a controversial topic is increased in length by 50% by the addition of threatening material presented throughout the communication, the supplemented message tends to result in (1) a higher proportion of statements indicative of anxiety, (2) a higher proportion of statements criticizing the content and presentation of the communication, and (3) a lower proportion of statements paraphrasing communication recommendations. These results are interpreted by Janis and Terwilliger (1962) "...as supporting the following general hypothesis: When a relatively high level of fear is induced by the warnings presented in a persuasive communication, the recipients will become motivated to develop psychological resistances to the communicator's arguments, conclusions, and recommendations."

The senior author of this study is also the senior author of a previous study (Janis and Milholland, 1954) pertaining to communication effectiveness in relation to emotional arousal. That previous study is not reviewed in this report as a separate study because of the desire here to restrict this review to those studies subsequent to the Janis and Feshbach study (1953) which obtained a measure of experimental effect. While the Janis and Milholland study involved no such measure, the results pertaining to retention would appear to me to have implications for the interpretation of the Janis and Terwilliger results.

The Janis and Milholland (1954) study was designed to test the hypothesis that arousal reduces the effectiveness of a communication by inhibiting retention of the content of the communication. They measured the amount of immediate recall of well-educated adults exposed to either a High Threat communication (N=24) or a Low Threat communication (N=22). The communication consisted of printed material excerpted from a short article on dental hygiene which appeared in the *Journal of the American Dental Association*, and which was based on the lecture material previously used by Janis and Feshbach (1953). After reading the article in an individual session, the subject was asked to write down as much of the material as he could remember.

The results revealed no significant difference between the two groups with respect to over-all recall. However, when 7 supplementary statements were divided into two types - threatening and explanatory - there emerged a significant interaction between threat level and statement type: the High Threat subjects recalled more of the threatening statements but fewer of the explanatory statements. In the Janis and Terwilliger study (1962) High Threat subjects gave a higher proportion of statements criticizing the communication, but a lower proportion of statements paraphrasing communication recommendations.

The parallelism implied in the two preceding sentences may seem strained. However, the results of Janis and Milholland (1954) and Janis and Terwilliger (1962) seem to me to be sufficiently similar as to encourage an attempt to subsume both results under a single explanation. One is encouraged in this attempt by the fact that the two studies utilized similar subjects and procedures (individual sessions, written communications, unstructured response, etc.).

Possibly the simplest explanation would be in terms of the differential effect of arousal level on (1) attitude toward the arousing material, and (2) retention of informative material, such as factually-based explanations, arguments, conclusions, and recommendations. Such informative material in each study was less frequently recalled by subjects in the High Threat group. This apparent similarity in the results of the two studies was not mentioned by Janis and Terwilliger (1962), and their only reference to Janis and Milholland (1954) occurred in the first sentence, which listed the progenitors of the Janis and Terwilliger study.

In view of the apparent effect of arousal on immediate recall, and in view of the evidence cited by Janis and Milholland (1954, page 75) of such an effect and of its temporary nature, it is surprising to find in all three of the studies in which Janis is senior author (Janis and Feshbach, 1953; Janis and Milholland, 1954; Janis and Terwilliger, 1962) the absence of a single statement which qualifies temporally the defensive avoidance (or resistance) hypothesis.

1.17 Snider (1962)

The subjects in the Snider study (1962) consisted of about 1500 junior and senior high school students. The investigator used three experimental variables: threat (of getting lung cancer), occurrence (of lung cancer in the general population), and defense (against getting lung cancer). Messages reflecting these three variables were included in a written communication on lung cancer in relation to smoking, with all messages occurring in the order of threat, occurrence, and defense, and with all communications containing the same information and recommendations. Since each of the three variables was associated with a High and Low message, there were 8 experimental groups. It was predicted that

High Threat and High Occurrence would result in a greater induction of arousal, while High Defense would result in a greater reduction of arousal.

At the end of the threat and occurrence messages, the student responded to a scale intended to measure anxiety. Then after the defense message was read, the student (1) answered 5 questions pertaining to his appraisal of the communication (accuracy, interest, informativeness, pleasantness, and importance), (2) responded to a scale intended to measure anxiety, (3) responded to an attitudinal scale which had also been administered two weeks previously, and (4) indicated whether he would like to help the American Cancer Society in its educational program. The variables utilized in the Snider study are indicated in Table 1.15.

TABLE 1.15

VARIABLES UTILIZED IN THE SNIDER STUDY (1962)

No.	Variable	No. of Items	2 Weeks Before	During	Just After
1	threat (as measured by anxiety)	7		x	
2	occurrence (as measured by anxiety)	7		x	
3	defense (as measured by anxiety)	7			x
4	appraisal of communication	5			x
5	attitude change (toward smoking)	20	x		x
6	verbal indication of future behavior	1			x

An analysis of variance of the anxiety scores revealed a significant threat effect, a significant occurrence effect, and a significant defense effect, with each in the predicted direction.

Students who received the High Threat, relative to those who received the Low Threat, appraised the communication as being more accurate, more interesting, more informative, and less pleasant, with all differences significant. The accuracy, interest, and informativeness items were all positively and significantly correlated, but these three items were not correlated significantly with the pleasantness item.

An analysis of variance of the attitude change scores revealed a significant threat effect with High Threat being associated with the communication-recommended direction, while the occurrence and defense effects were not significant. The combined experimental groups differed significantly from a control group with respect to attitude change in the communication-recommended direction.

The experimental groups did not differ significantly with respect to verbal indication of future behavior, as measured by the number of students volunteering to help the American Cancer Society.

The above summary of Snider's results does not indicate the probability level associated with each significant result; these levels range from p=.025 to p=.001.

Snider (1962) also compared High and Low Threat using non-smokers only, then smokers only. High Threat non-smokers showed significantly higher anxiety than did Low Threat non-smokers, and High Threat non-smokers showed significantly greater attitude change in the recommended direction than did Low Threat nonsmokers. Even though High Threat smokers did not differ significantly from Low Threat smokers with respect to anxiety, High Threat smokers showed significantly greater attitude change in the recommended direction than did Low Threat smokers. As might be expected, smokers as a group were significantly higher than nonsmokers as a group with respect to anxiety.

Following the arguments of Janis and Feshbach, Snider hypothesized that the Low Threat condition would show the greatest attitude change in the direction advocated by the communication. The results were just the opposite and highly significant. There were three three parts to the attitude questionnaire, and on each part the change in attitude in the direction supported by the communication was significantly higher for the High Threat group.

Here, then, we have the rather convincing demonstration that higher arousal during exposure to a communication can enhance the effectiveness of that communication.

1. 18 Summary

The three studies conducted during the 1935-1940 period wil not be included here; a summary of the results of these three studies has been given at the end of Section 1. 11. The following material, then, pertains only to the remaining six studies.

The preceding review has considered each of these six studies separately. However, this summary will proceed topically across studies in the expectation that such an organization will allow a more clear emergence of the implications which these past studies have for the current study. The topics are organized under (A) Method, (B) Results, and (C) Interpretations.

A. Method

1. Subjects

Five of the six studies used students, ranging from junior high school through college. In addition to the college students used by Janis and Terwilliger (1962), a few adults were included. Moltz and Thistelthwaite (1955) used newly-inducted Air Force recruits.

2. Treatments

Each study utilized treatments intended to induce two levels (three levels in the case of Janis and Feshbach) of emotional involvement (anxiety, fear, arousal, etc)

Each study utilized mass media communications to induce the treatment effect. The first four studies utilized slide-tape lectures, the last two printed materials. One study (Janis and Terwilliger, 1962) presented the communications individually to subjects, whereas the other studies administered the communications to groups of about class size. The communications pertained to dental hygiene (the first three studies), automobile safety belts (fourth study), or smoking and lung cancer (last two studies). The comparison forms of the communications were intended to contain the same information and advocacy; they were intended to differ only with respect to the degree to which the emotions of the subject would be aroused (and reduced in the case of two studies). Thus, each study attempted to manipulate arousal solely by means of the communication. It is not at all clear the extent to which the authors assume that in order for arousal (or arousal-reduction or arousal-consonant) to be related to the effectiveness of a communication, the arousal must be induced by the communication. This assumption was probably quite general, since not one of these studies attempted to induce arousal by means other than the communication, or discussed this possibility. Further, each study emphasized the relation between communication content and communication effectiveness.

3. Experimental Variables

Each study utilized as the primary experimental variable one which in this report is sometimes referred to as emotional involvement. This variable has been referred to in the six studies by some 20 different terms (anxiety, fear, arousal, etc.) with each study using several such terms. To assume that each study used the same experimental variable is a simplification, of course, but none of these studies attempts to make a clear distinction between its experimental variable and those used in the other studies. In the absence of any standardized operational procedure for defining this general variable, it will be assumed that each study, in a gross sense, was utilizing the same experimental variable. In referring to this general variable, this report will utilize the most commonly-used terms employed in the six studies being reviewed.

Moltz and Thistlethwaite (1955) utilized another primary experimental variable - anxiety reduction. Snider (1962) utilized two anxiety-inducing variables and one anxiety-reducing variable.

All experimental measures were obtained only at discrete points in time. While the interest in each study was in the degree of arousal which occurred during exposure to the communication, four of the studies obtained arousal measures only after completion of exposure The two exceptions were Snider (1962), who interrupted the exposure once to obtain an additional measure of arousal, and Janis and Terwilliger (1962), who interrupted the presentation 15 times. The first three studies obtained a pre-communication measure of anxiety, while the last three did not.

All experimental variables in these studies were measured verbally.

4. Effectiveness Variables

Each study utilized at least one effectiveness variable as a dependent variable, with the first three studies utilizing more than one effectiveness variable. The primary effectiveness variable in the first three studies was verbal conformity to communication-recommended dental practices, as measured one week (two weeks in the Goldstein study) after the communication. A secondary effectiveness variable in these three studies was information retention, as measured immediately after (as well as two weeks after in the Goldstein study) the communication. Janis and Feshbach (1953) utilized two additional secondary effectiveness variables: change in communication-recommended belief, and change in communication-recommended attitude, with the change based on measures obtained one week before the communication and one week after.

Each of the last three studies utilized only one effectiveness variable: attitude change. This variable was based on measures obtained one month before and immediately after the communication in the Berkowitz and Cottingham study, immediately before and immediately after in the Janis and Terwilliger study, and two weeks before and immediately after in the Snider study.

All effectiveness variables in these studies were measured verbally.

5. Other Variables

Five of the studies used additional variables, and all of these other variables are listed in this section. Janis and Feshbach (1953) utilized a variable called "appraisal of communication." Goldstein (1959) utilized two levels of a personality variable. Berkowitz and Cottingham (1960) utilized three variables: "appraisal of interest value," "aggression toward communicator," and "relevance of communication." Janis and Terwilliger (1962) obtained two measures of resistance (criticism of communication, and paraphrasing of communication recommendational). Snider (1962) utilized variables called "appraisal of communication," and "verbal indication of future behavior."

All of the variables in this section were measured verbally, as was the case with all variables in the two preceding sections. Thus, all variables in these studies were measured verbally; Janis and Terwilliger (1962) utilized oral measures, whereas the other five studies utilized questionnaire measures.

B. Results

A given result is sometimes interpreted in this report as being associated with a null hypothesis probability level which is different from that given in the published study. It is left to the reader to decide which determination is correct. For the purpose of this review (but not for the research undertaken in this report) the .05 probability level is taken as the significance level, as mentioned at the beginning of this report, and each probability level throughout this report is two tailed.

1. Differential Arousal

The term "differential arousal" pertains to the difference between the Strong (High) and Minimal (Low) groups with respect to the experimental variables. The experimental effects mentioned in the next paragraph were all in the predicted direction.

Moltz and Thistlethwaite (1955) showed a significant anxiety-reducing effect. Janis and Terwilliger (1962) demonstrated a significant anxiety effect. Snider (1962) demonstrated significant anxiety-reducing effects. The remaining three studies failed to demonstrate a significant experimental effect.

2. Differential Effectiveness

The term "differential effectiveness" pertains to the difference between the Strong (High) and Minimal (Low) groups with respect to the effectiveness measures.

In the Berkowitz and Cottingham study, the Strong group was significantly higher than the Minimal group with respect to attitude change in the communication-recommended direction. In the Snider study, the High Threat group was significantly higher than the Low Threat group with respect to attitude change in the communication-recommended direction. The remaining four studies failed to demonstrate any significant results involving the effectiveness variables.

3. Other Results

The results in this section constitute all the statistically significant results involving the variables listed above under A5(Other Variables).

In the Berkowitz and Cottingham study, the three Relevance subgroups within the Strong group differed significantly with respect to attitude change, with relevance and attitude change showing a negative relation; however, because of the way in which Berkowitz and Cottingham (1960) present their data on appraisal of interest value, it is not possible to tell whether the two groups differed significantly with respect to this variable. In the Janis and Terwilliger study, the High Threat group appeared to be significantly higher than the Low Threat group with respect to criticism of the communication, and significantly lower with respect to paraphrasing of communication recommendations. In the Snider study, the High Threat group, relative to the Low Threat group, appraised the communication as being significantly more accurate, more interesting, more informative, and less pleasant.

C. Interpretations

The interpretations given in this section are those given by the authors of the six studies which are related to the Janis-Feshbach hypothesis.

Janis and Feshbach (1953), on the basis of their defensive avoidance hypothesis, predicted that a communication which induces a higher level of anxiety will be less effective than one which induces a lower level of anxiety.

The preceding review covered not only the study of Janis and Feshbach (1953) but also five subsequent studies, each of which (1) utilized two forms of a communication intended to differ only with respect to degree of anxiety aroused in the communicatee, and (2) obtained at least one measure of communication effectiveness. Of these five subsequent studies, not one gave results in support of the Janis-Feshbach hypothesis: each of three studies revealed no significant difference between the High and Low groups with respect to effectiveness, while each of the remaining two studies revealed a significant difference in effectiveness in favor of the High group, a result opposed to the Janis-Feshbach hypothesis. Yet oddly enough, not a single one of these five studies interpreted the results as failing to support the Janis-Feshbach. Quite the contrary, the authors tended to interpret their results as supporting the Janis-Feshbach hypothesis - namely, that a high anxiety communication is less effective. The arguments by which this reversal was achieved are sufficiently interesting to be noted briefly here.

Moltz and Thistelthwaite (1955), failing to obtain a differential arousal, and assuming that Janis and Feshbach (1953) did, explore the possibility that the assumed discrepancy between their results and those of Janis and Feshbach might be due to the change in the lectures. Recall that in the Janis-Feshbach versions, the recommendations concerning the proper dental hygiene practices were interwoven with threatening material, whereas in the Moltz and Thistelthwaite versions, the recommendations were placed at the end. Moltz and Thistlethwaite suggest the possibility that interweaving recommendations and threatening material may actually make the communication more anxiety arousing. Thus, they imply that they may have failed to induce a differential arousal because their versions did not differ as much as did the Janis and Feshbach versions in their capability to induce arousal.

Goldstein (1959, page 251) notes that his study "... leaves the conclusions of Janis and Feshbach largely unchanged. A minimal fear appeal is still most effective in eliciting acceptance of propaganda."

Berkowitz and Cottingham (1960), obtaining significant results opposite to those predicted from the Janis-Feshbach hypothesis, use the following argument to bring their results in line with that hypothesis. Berkowitz and Cottingham suggest that the Janis-Feshbach communication was interesting to even the Minimal group, whereas their communication "... may be low in intrinsic interest for the audience." On this basis Berkowitz and Cottingham argue that their Strong appeal was probably equal in its anxiety-arousing capability to that of Janis and Feshbach's Minimal appeal, and that since Janis and Feshbach's Minimal appeal was effective, so also should Berkowitz and Cottingham's Strong appeal. However, this argument fails to explain why the appeal which then automatically becomes the strongest (Janis and Feshbach's Strong) and the appeal which then automatically becomes the weakest (Berkowitz and Cottingham's Minimal) emerge as the least effective in their respective studies, a problem which is raised because of the implication in each of these studies that communication effectiveness is a single-valued function of arousal. Further, the authors fail to suggest why a topic related to teeth is intrinsically more interesting for American juveniles than a topic related to automobiles.

Janis and Terwilliger (1962) obtained a differential effectiveness in the predicted direction, but the result was not significant. The authors argue, however, when this result is evaluated along with the results of the Janis and Feshbach study (see Table 1.7 for a summary of these results), the non-significant result in the Janis and Terwilliger study nevertheless "... contributes to the general conclusion that when a strong threat appeal evokes a high degree of affective disturbance, it will tend to be less effective in inducing attitude change than a milder threat appeal."

Snider (1962) also attempts to reconcile his results with the Janis-Feshbach hypothesis. His task is made the more difficult because his results, opposite to those predicted from the Janis-Feshbach hypothesis, are so highly significant. As one possiblilty for resolving the discrepancy, Snider follows the argument of Berkowitz and Cottingham (1960) - namely, that Snider's High Threat condition is really equal in its anxiety-inducing capability to Janis and Feshbach's Minimal condition.

This comes as a surprise, in view of the fact that Snider's High Threat appeal included such statements as "A million boys and girls like you will die of lung cancer before you reach age 40," "If you smoke you are gambling with your life," "Soon this lung [picture] will wither away and die. If you smoke, this is what might happen to your lung," and "If you smoke, there is one chance in 5 that you will get lung cancer during the peak of your life." On the other hand, the most threatening material in Janis and Feshbach's Minimal appeal were such statements as "Recent surveys in many communities show that approximately 90% of the high school students examined were in need of dental care," "In this x-ray picture we can see small cavities beginning to form," and "Remember, you cannot expect to have good health unless you also have good teeth." While it seems to me that Snider's High Threat appeal would be expected to be more anxiety inducing, Snider suggests that the reverse may in fact be the case. He argues that a person who is presented a minor threat if he fails to conform may in fact be more prone to defensive avoidance. Snider concludes that

> "If this is so, it may explain why a high fear group in the present study behaves like a low fear group in another study where the penalty for nonconformity is less severe. This suggests that while the general principles observed by Janis and Feshbach and others may apply to different content areas, the way in which 'high' and 'low' fear are defined may be unique to each content area."

1.2 Research on Learning and Remembering

Since films constitute one medium of mass communications, the preceding review covered an area which would seem to be obviously relevant to the current study. One assumption (sometimes implicit) of the preceding studies was that the learning and remembering of communication information constitute two factors which influence communication effectiveness. This assumption is made also in the current study.

The current study utilizes information retention as its only measure of effectiveness. Frequently, as in some of the studies already reviewed, short-term retention is taken as an indicator of learning, whereas long-term retention is taken as in indicator of remembering. This is probably too great a simplification; it is probably more appropriate to think of retention as reflecting both learning and remembering. In view of the focus on retention in the current study, it seems appropriate to review related studies on learning and remembering. The full breadth of this area cannot be reviewed here, and arbitrary restrictions have been imposed. As in the previous section, the review will be restricted to recent studies involving human subjects.

In recent years renewed interest has been shown in the two concepts of the orienting response and of arousal, and both areas appear to have implications for information retention. Selected examples of studies in these research areas will now be reviewed.

1.21 Learning and the Orienting Response

The concept of the orienting response (OR) appears to be pertinent to an understanding of the relation between arousal and retention. The discussion here of this concept will rely heavily on recent work performed at the University of California, Los Angeles (hereafter referred to as UCLA) by Maltzman and his colleagues, with much of this work summarized by Maltzman and Raskin (1965).

The orienting response may be considered as a nonspecific respondent to any temporal change in the quality or intensity of stimulation. As with other respondents, it decreases with repetition of the same stimulus. The OR may be measured by pupillary dilation, cephalic vasodilation, peripheral vasoconstriction, the GSR, and reduction in the power of the alpha rhythm. As might be expected, these functions are correlated in time, and are responsive to changes in emotional states (Hess and Polt, 1960) and mental activity (Hess and Polt, 1964). The OR may occur at the same time as an operant (Maltzman and Raskin, 1965), creating difficulties in distinguishing the two when their manifestations are similar. OR level is susceptable to a number of antecedent conditions.

The orienting response facilitates the reception of stimuli and the lowering of sensory thresholds (Sokolov, 1963); the results of Eriksen and Wechsler (1955) are consistent with this interpretation. At least a portion of the enhanced receptivity appears to have its locus at the periphery. Further, the OR can be conditioned to previously-neutral stimuli, including complex, meaningful stimuli. As with any other response, there are wide individual differences in the orienting response.

Maltzman and Raskin (1965) review a number of studies which demonstrate the relationship of the OR to learning. In one study a subject who gave a larger OR to an unconditioned stimulus (white noise) tended to show superior conditioning of the OR to a previously-neutral test word, as well as greater semantic generalization of the test word to other words presented during the trial, and associated with the test word on the Kent-Rosanoff list. Semantic generalization was measured verbally after the experiment. Thus, high orienters show not only superior conditioning but also greater awareness of the experimental contingencies. In two other studies, subjects classified as high orienters on the basis of the magnitude of their OR to a standard stimulus were superior to low orienters in easy and difficult paired-associates learning. These studies are described in greater detail in Section 4. 3.

Maltzman and Raskin suggest that performance is more closely related functionally to the OR than to anxiety or emotionally-based drive. They conceive of the orienting response as enhancing discriminative ability, rather than as energizing behavior. They advance the notion that the OR is a fundamental determiner of learning by acting as a reinforcer without necessarily involving drive or drive reduction - that is, the OR itself supplies the reinforcement necessary for learning. Finally, they note the possibility that the OR may not be a single nonspecific response, but a class of responses, each a function of antecedent conditions, and that this possibility poses a problem to the concept of the OR.

While arousal is not equivalent to OR, a study by Rosenstein (1960) tends to support the Maltzman and Raskin postulation that performance is more closely related to OR than to anxiety. As his measure of arousal, Rosenstein utilized PSI (Palmar Sweat Indices). Measures were obtained immediately before the experimental stress conditions were imposed, then again immediately after nonsense syllables were learned under one of the stress conditions. Subjects who showed a greater PSI increment learned the material in fewer trials and with fewer errors, a finding common to five experimental conditions. However, learning was not related to anxiety, as measured by the Mandler-Sarason Test Anxiety Questionnaire (Mandler and Sarason, 1952).

Interest here in the OR is primarily in its relationship to learning and not in the biological basis of this relationship. Nevertheless, it is not inappropriate to consider the process by which a rapid change in stimulation effects not only the OR but also enhances learning.

The reticular formation might very well be involved in such a process. There is evidence that the ascending extensions of the reticular formation constitute a nonspecific sensory system which is activated by specific sensory inputs and which activates widespread areas of the cortex (Lindsley, 1957). This reticular arousal system (RAS) effects the general arousal function, which is assumed to affect both attention and perception.

Under normal, slowly-changing stimulation, the cortex sends anticipatory impulses to the RAS of an inhibitory nature. However, if the RAS is activated by a surprise stimulus before the cortex is able to transmit inhibitory corticoreticular impulses, the RAS effects a general arousal of widespread portions of the cortex. In effect, the cortex is caught off guard (Berlyne, 1960). At the same time, activity in the descending extensions of the reticular formation may effect a change in skin resistance by way of the autonomic nervous system.

Whatever the mechanism, a rapid change in stimulation will tend to result in cortical and autonomic arousal which are likely to extent over some period of time. It is assumed here that this arousal is related to enhanced learning and memory. The function of this arousal over time is determined by the damping characteristics of the system. Whatever the damping characteristics, hunting (oscillatory arousal) would be expected by virtue of the fact that (1) the corticoreticular system involves negative feedback, and (2) the input to this system is a step function (sudden change in stimulation).

It may be argued that the OR is related only to learning, especially its attentional and perceptual aspects, and that the OR is not related directly to remembering. Even if this were true, factors which enhance learning are likely to be manifested in measures of retention.

1. 22 Remembering and Arousal

The extent to which retention is influenced by arousal is dramatically revealed by a series of recent studies performed at the University of Michigan (hereafter referred to as UM) by Walker (1962) and his colleagues (Kleinsmith and Kaplan, 1963; Kleinsmith and Kaplan, 1964; Kleinsmith, Kaplan, and Tarte, 1963).

In these studies arousal was measured physiologically, using GSR as the measure of arousal. Each of these studies shows a strong relationship between arousal during learning and subsequent retention, such that high arousal, relative to low arousal, is associated with poor immediate recall and good delayed recall. The phenomenon emerged whether the analyses were between-subjects (Kleinsmith, Kaplan, and Tarte, 1963) or within-subjects (Kleinsmith and Kaplan, 1963, 1964), whether arousal was manipulated experimentally (Kleinsmith and Kaplan, 1963) or not (Kleinsmith and Kaplan, 1964; Kleinsmith, Kaplan, and Tarte, 1963), or whether the stimulus materials were meaningful (Kleinsmith and Kaplan, 1963; Kleinsmith, Kaplan, and Tarte, 1963) or meaningless (Kleinsmith and Kaplan, 1964). This latter indicates that while arousal may be related to meaningfulness, arousal is the determining variable, not meaningfulness, as was assumed in the past (Edwards and English, 1939; Williams, 1926).

These results have been interpreted by the experimenters in terms of neural consolidation. They argue that learning under high arousal will result in a consolidation process which engages a greater number of neural cells and which perseverates for a longer period of time. Under these conditions (1) immediate recall is poor because of the relative unavailability of actively consolidating traces, and (2) delayed recall is good because of the stronger consolidation of these neural traces.

The UM results suggest that a high arousal communication is likely to be more effective in the long run than a low arousal communication, if effectiveness is measured by retention of communication content.

This postulation is essentially opposite to that of the studies summarized in Section 1. 18, and is also different from another common orientation, which holds that a moderate level of arousal, relative to a low or high level, is most favorable to learning and retention (Beam, 1955; Berlyne, 1960; Berry, 1962; Hebb, 1949; Luborsky, Blinder, and Mackworth, 1963). However, there are so many exceptions to the superiority of moderate arousal for learning as to make its universality unlikely (Jackson and Strattner, 1964; Maltzman, Eisman, and Morrisett, 1961). Development of this area has been hampered by the practice in some studies of defining level of arousal on the basis of a subjective evaluation of the arousal material. Under these conditions it is highly unlikely that different conditions defined as "high" arousal by different investigators do in fact mediate the same level of arousal.

Further, the UM results emphasize the importance of the point in time at which the retention measure is obtained. The apparent superiority of moderate arousal may be due to a procedural artifact in those studies which obtain only an immediate, or short-term, measure of retention. In at least one such case (Berry, 1962), the superiority of moderate arousal was subsequently shown (Kleinsmith, Kaplan, and Tarte, 1963) to be restricted to short-term measures only.

The UM results also emphasize the importance of identifying the amplitude of the GSR, rather than simply its presence. Failure to distinguish among GSRs on the basis of amplitude has the effect of treating as similar each learning unit (item, frame, stimulus, etc.) whose retention is later tested. This is true irrespective of the point in time at which the retention measure is obtained. This failure to identify a GSR during learning on the basis of its amplitude could be the basis for the equivocal results obtained by Becker (1964).

Becker (1964) recorded the skin resistance of fifth-grade students during individual viewing of a film. The student was tested on the factual content of the film both immediately before and after exposure. GSRs were identified simply as "changes in resistance." The total number of GSRs for each student was found not to be significantly related to information change or information retention.

Becker also determined whether each student gave at least one GSR in the 30-second interval in the film in which appeared information pertaining to the questionnaire. If the information for an item was given in more than one section of the film, all such sections were treated as one. It was found that a student who showed a GSR during a particular section of the film tended to learn (incorrect on pretest, correct on post-test) the information presented in that section, but on a second film the opposite result emerged. When the total number of GSRs was determined for each student, and the high learners compared with the low learners, the high learners gave a greater number of GSRs, though the difference between the high and low learners was not significant.

Becker also obtained an overt indication by the student at those points in the film which the student found interesting, with the expectation by the author that the student would show a high GSR at those points in the film he found interesting. Actually just the opposite relation emerged, leading Becker to the assumption that GSR was indicative of tension, not interest. This assumption tended to receive support from the fact that the appearance in the film of an authority figure (in this case, a teacher) was associated with a large number of GSRs.

In view of the fact that no reliable relation was found between GSR and retention, Becker concluded that tension arousal during film viewing has little to do with the new knowledge a student will gain from that viewing.

The interaction between arousal level and retention interval with respect to information retained is often overlooked. For instance, this interaction existed, but was not noted, in a study by Schlesinger, Fischer, and Cohen (1965). After an initial pilot study, these experimenters exposed two groups of driver education students to two self-instructional programs. One group was exposed to a program intended to induce a high level of arousal, the other group to a program intended to induce a low level of arousal. Each program occupied two sessions on two successive days. After the second session, the students responded to two questionnaires: the first yielded a verbal measure of arousal to the program, while the second yielded a measure of amount learned. As expected, there was a significant difference between the groups in arousal, and in the direction intended. The authors failed to find, however, a significant difference between the two groups in retention.

The aspect of this study which is of interest to the present discussion is the fact that an interaction emerged between arousal level and retention interval: low arousal resulted in higher immediate retention, whereas high arousal resulted in higher delayed (1-month) retention. This would be predicted on the basis of the arousal-interval interaction hypothesis, though it is not possible to test the significance level of the interaction from the data given by Schlesinger, Fischer, and Cohen (1965).

If it is true that greater long-term retention results from increased autonomic activity at the time of learning, then detection of the effect of experimentally-induced arousal requires the criterion measure to be obtained at a delayed point in time if groups are being compared, or at two points in time if the subject is acting as his own control. Thus two learning conditions may show no difference in immediate retention, but if one condition is associated with greater autonomic arousal, one would predict that that condition would be associated with superior delayed retention.

This line of thinking receives indirect support from the study of Krumboltz and Weisman (1962). Three groups of undergraduate students were exposed to the same instructional program. The "written response" group was instructed to write a response to each blank before turning the page to the correct answer. The "correct response" group was instructed to mentally compose a response to each blank before turning the page to the correct answer. The "reading only" group was instructed only to read the program, and in this case the correct answers appeared in the blanks. On the day following the completion of the program, a criterion test was administered to each subject, and two weeks later an alternate form of the criterion test was administered. The "written response" group showed a significantly greater retention than both the "correct response" group and the "reading only" group on the second test, but not on the first. The latter two groups did not differ significantly from each other on either test.

If one assumes that the act of responding overtly is associated with increased autonomic activity, then the Krumboltz and Weisman results are compatible with the hypothesis that autonomic arousal during learning results in superior long-term retention. Their results suggest that other studies which compared overt vs. covert response in programed instruction might have found a delayed superiority for the overt response had a delayed retention measure been obtained. In general, however, it is assumed that overt response is not superior to covert response with respect to retention (Alter and Silverman, 1962; Evans, Glaser, and Homme, 1960; Feldhusen and Birt, 1962; Goldbeck, Campbell, and Llewellyn, 1960; Kaess and Zeaman, 1960; Kanner and Sulzer, 1961; Roe, 1962; Stolurow and Walker, 1962).

McGuire (1961) also found no significant difference in short-term retention between overt and covert responding when material was presented at a slow rate, but overt responding was inferior when the rate of presentation was fast. This finding is compatible with the UM orientation if one accepts the assumption that in comparison to covert responding, overt responding under the fast rate of presentation was associated with a higher degree of arousal. This assumption seems reasonable in the light of the conditions associated with the fast rate: slides were projected at 2-second intervals, and the subject made an attempt to associate the mechanical part and name shown in the slide, as well as write down the name. The UM hypothesis would lead to the prediction that the inferiority of the overt response would be short-lived, and that a fast rate would favor the overt response for long-term retention. Unfortunately this prediction cannot be tested with McGuire's data because a delayed retention measure was not obtained.

The consolidation process is assumed to act over time in such a manner that the neural representation of the learning material becomes increasingly fixed. Consolidation may take place over a period of 15 minutes (Pauling, 1961), an hour (Pribram, 1964), a month (Brady, 1952), or conceivably a lifetime (Glickman, 1961). Whatever the length of the consolidation period, it is assumed that the consolidation process is liable to interference from external sources. It is on this basis that perseveration theory is used to account for retrograde amnesia. The idea dates back to the turn of the century, at which time it was suggested that the fixing of an impression involves a physiological process which takes a considerable length of time, and that this process could be interrupted by shock, extreme fatique, or excitement (Burnham, 1903). Since then it has been learned that retrograde amnesia also occurs upon application of electroconvulsive shock, anoxia, increased or decreased temperature, anesthesia, and brain stimulation (Glickman, 1958; Pearlman, Sharpless, and Jarvik, 1961).

A suggested basis for consolidation is the reverberatory hypothesis (Hebb, 1949), which considers reverberatory electrical activity within a group of cells as the neural representation of the learning material. In support of this hypothesis is the finding that electrical activity in isolated cells may last for 30 minutes or longer (Burns, 1954). This conceptualization is consistent with, but is not required by, the UM results. In any case, the primary interest here in the UM results is their implications for communication effectiveness, not for the biological basis of memory.

1.3 Implications for Current Study

Two research areas have been reviewed: (1) communication effectiveness and arousal, and (2) learning and remembering. Both areas have implications for the current study. The two research areas differ with respect to the type of theories employed, as well as with respect to methodology. Since it seems appropriate to separate theory from methodology, the implications of each will be discussed separately.

In this section and throughout this report, the term "studies of Section 1.1" include only the six studies summarized in Section 1.18.

1.31 Theoretical Implications

This section will consider first the implications of the research reviewed in Section 1.1 and then the implications of the research in Section 1.2.

First, an attempt will be made to classify the six communication studies according to their advocacy of the optimal degree of arousal, optimal in the sense that it leads to greater communication effectiveness. At the risk of oversimplification, it seems to me that the six studies can be classified in three categories: arousal minimum, arousal reduction, and arousal consonant.

The arousal-minimum position is held by persons who state that, other things being equal, that communication will be most effective which results in the least arousal. The arousal-reduction position is held by persons who state that, other things being equal, that communication will be most effective which results in the greatest reduction in arousal. The arousal-consonant position is held by persons who state that, other things being equal, that communication will be most effective which results in that level of arousal at which learning is predispositionally optimal for the subject.

According to this schema, it would appear that the positions of Janis and Feshbach (1953), Berkowitz and Cottingham (1960), and Janis and Terwilliger (1962) should be classified as arousal minimum, the position of Moltz and Thistlethwaite (1955) and Snider (1962) as arousal reduction, and the position of Goldstein as arousal consonant. Berkowitz and Cottingham (1960) may not appear to be taking the arousal-minimum position, for in their study they predicted, allegedly, that their Strong appeal would be more effective. However, this prediction was based on the argument that their particular Strong appeal would be more interesting, and that it would be by virtue of this enhanced interest value that such an appeal would be more effective. Other things being equal (interest, relevance, etc.), Berkowitz and Cottingham would predict that that communication which results in the least arousal will be most effective.

An even simpler classification may be suggested, one consisting of only two categories: level and change. The level orientation is held by persons who state that, other things being equal, that communication will be most effective which results in that level specified by the advocate of this position. The change position is held by persons who state that, other things being equal, that communication will be most effective which results in that direction of change specified by the advocate of this position.

According to this second schema, it would appear that the positions of Janis and Feshbach (1953), Goldstein (1959), Berkowitz and Cottingham (1960), and Janis and Terwilliger (1962) are level oriented, while the positions of Moltz and Thistlethwaite (1955) and Snider (1962) are change oriented.

So much for schematization of the position taken by these six studies. It is important to note that the schema pertains to position, not evidence, for it is recalled that none of these positions obtained critical support from the results of these six studies.

Next, we turn to the implications which these six studies have for the current study. These implications emerge more clearly if they are limited to those studies which demonstrated a differential effectiveness between treatment groups. The Berkowitz and Cottingham study and the Snider study were the only ones of the six which qualify. In both cases the effectiveness measure (attitude change) was related positively to high arousal and, under the assumption that both groups started from the same arousal level, to arousal induction, but was not related to arousal reduction. Thus, the interpretation given here of the composite results of a decade of mass communication studies in the area of communication effectiveness in relation to arousal is that communication-mediated attitude change is related positively to high arousal or arousal induction.

It is unfortunate that these two substantial results pertain to attitude change, not information retention. This fact makes it difficult to draw implications for the current study, which is restricted to information retention. While attitude change and information retention must no doubt be related, the nature of the relationship is apparently complex, as the Janis and Milholland (1954) and Janis and Terwilliger (1962) studies show (see Section 1. 16). These two studies seemed to suggest that the immediate effect of a high arousal communication is to facilitate emotionally-toned attitude change and to inhibit information retention.

It is also unfortunate that arousal was measured in these two studies, as well as in the other studies in Section 1. 1, by a method which does not allow a clear distinction to be made between arousal level and arousal change. The subject was asked in effect to indicate the way he felt during the communication. Invariably, the question pertained to level, but any report of arousal above the pre-communication level also implies an arousal induction. Hence, high arousal is indistinguishable from arousal induction.

Even though the current study falls into the class of studies in Section 1. 1 - that is, the current study is concerned with communication effectiveness in relation to arousal - the studies of Section 1. 2 probably have more direct implications for the current study. The two limitations listed in the two preceding paragraphs are much less applicable to the Section 1. 2 studies, which are concerned more with information retention, not attitude change, and which seem to allow a distinction between arousal level and arousal change.

The UCLA studies suggest that the magnitude of the GSR is the important factor in learning, irrespective of the level at which the GSR occurred. The basic procedure involved the specification of OR on the basis of a single GSR measure (resistance change) following an adaptation period (in effect, 16 adaptation words, each followed by relative silence). Under this condition it would be expected that the change started essentially from a basal level. However, the UCLA studies obtained comparable results whether the GSR was measured under the condition above or after the first word following 3 minutes of background noise (Maltzman and Raskin, 1965). Under this latter condition, it is unlikely that the change began from the basal level. Hence, the important parameter seems to be the extent of the resistance change and not the resistance level before or after the change.

The UM studies use the term "arousal level", but despite this fact, it is quite clear from their procedure that in at least two of the studies (Kleinsmith and Kaplan, 1963, 1964) the term is used in the same sense as the term "arousal change" is used in this report. Each of these studies involved a single learning session consisting of 6 to 8 sequences, with each sequence consisting of 4 presentations: (1) the stimulus member of a paired-associate, (2) both stimulus and response members of the pair, (3) 4 colors, and (4) another 4 colors. Each presentation lasted 4 seconds, during which time the subject identified out loud the material being presented. The recall session was similar except that (2), above, was omitted; the subject called out the response member when (1), above, was presented, guessing if necessary. The recall session followed the learning session by an interval ranging from two minutes to one week. A subject was run at only one recall interval.

An arousal change was taken as any detectable drop in skin resistance which occurred within 4 seconds of the presentation of (1), above. A larger drop was considered as "higher arousal," in the terminology of the UM studies. If the drop always started from the same (basal) resistance level, drop size and post-drop resistance level would be inversely related, and arousal change would be directly related to arousal level, as these terms are used in this report. However, the UM procedure would not appear to be such as to allow resistance to return to its basal level: during a learning session lasting only one or two minutes, the subject had to identify orally material presented every four seconds, with some of this material involving four elements. Under these conditions it is not likely that the subject's skin resistance ever returned to a basal level. Under this assumption, then, it appears that the arousal-induction hypothesis is more appropriate than the arousal-maximum hypothesis

In any case these two hypotheses are similar, since a high arousal level measured at any point in time, averaged over an interval, during learning is likely to imply arousal increments while the preceding learning material was presented. Thus, the results of a study are likely to be similar whether arousal is measured as a level or as a change. Evidence for this is given by the fact that the UM results were obtained whether arousal was measured as a level (Kleinsmith, Kaplan, and Tarte, 1963) or as a change (Kleinsmith and Kaplan, 1963, 1964).

The UCLA and UM studies differ in a number of ways. The UCLA studies obtain one GSR measure at the time the subject is exposed to the unconditioned (or standard) stimulus for the first time; the UM studies obtain GSR measures each time the subject is exposed to material to be learned. Under these conditions, the UCLA studies interpret GSR in terms of the orienting reflex; the UM studies interpret GSR in terms of arousal. The UCLA results indicate a positive relation between the OR and learning; the UM results indicate an interaction between arousal and retention interval, such that higher arousal is associated with better long-term memory. The UCLA studies explain the influence of the OR on learning in terms of increased discriminative ability; the UM studies explain the influence of arousal on remembering in terms of perseverative consolidation.

Are these differences amenable to a common discussion? Possibly, if a given GSR affects both learning and remembering. In discussing this possibility, it seems appropriate to relate GSR to continuous changes in skin resistance.

Continuous behavioral measures are not commonly used in psychology. Rather, behavioral measures are generally obtained during specific time intervals. The limits of this interval are usually determined by the investigator. For instance, the GSR is often measured during the brief interval following the presentation of a discrete stimulus, though in some cases the measurement of GSR is not restricted to a particular interval but is made at any point throughout a continuous record where the skin resistance change meets certain criteria with respect rate and/or amplitude.

Whatever the relationships of skin resistance change to learning and remembering, they would seem to be independent of the particular interval of time the experimenter chooses to investigate. If resistance change is associated with increased discriminative ability, or increased perseverative consolidation, or increased reticular activity, or decreased synaptic threshold, the processes involved presumably occur at times other than those of interest to the experimenter. And if rapid changes of skin resistance are associated with these processes, the association must also be independent of the environmental conditions antecedent to the resistance changes. For instance, if a resistance change occurs just before the experimenter presents the unconditioned stimulus, the change would not be taken as a manifestation of the OR, but according to the formulation suggested here, that resistance change would have the same influence on learning and remembering as a resistance change immediately following the unconditioned stimulus. If the second resistance change enhances discriminative ability, it is assumed here that the first will too, though the effect of the first change will not be detected because the effect is with respect to environmental events of no interest to the experimenter.

The argument being advanced here in no way vitiates the position that the OR must be defined in relation to the stimulus (antecedent) conditions. Rather, it is suggested that the S-R paradigm is not essential to an understanding of the relation between skin resistance change on the one hand, and learning and remembering on the other. A resistance change is a resistance change whether it is conceptualized as a response to a known or unknown stimulus, or as an adaptation of a servomechanism.

In line with this formulation, it should be possible to summarize the UCLA results, as well as the UM results, in relation to resistance change. The simplest summary is the following: resistance change reflects a process (consider in the singular for simplicity) which influences learning and remembering. How much of this influence should be attributed to an increase in discriminative ability, and how much to perseverative consolidation?

For the sake of efficiency here, and throughout this report, the word "initiation" will be used to mean the initiation of the process which presumably enhances discriminative ability (discrimination) and engram consolidation (consolidation), and the word "presentation" will be used to mean the presentation of the learning material or stimulus.

It should be possible to identify the effect of enhanced discriminative ability and engram consolidation. If initiation occurs before presentation, the enhancement of discrimination and consolidation would be proportional to the intensity of the process which remains at the time of presentation. On the other hand, if the learning material is introduced and removed before initiation, then discrimination would not be enhanced during the presentation period, and the enhancement of consolidation would be proportional to the engram representation which remains at the time of initiation. Even though both enhanced discrimination and consolidation would operate in the first case, whereas only enhanced consolidation would operate in the second case, it does not necessarily follow that the first condition will result in greater long-term retention. In order to make any prediction regarding which of the two conditions will result in the greater long-term retention, it would be necessary to have knowledge of three functions; (1) discrimination following initiation, (2) consolidation following initiation, and (3) engram representation following presentation. Since these functions are not known at this time, we must resort to an empirical determination of whether long-term retention is enhanced more by initiation followed by presentation, or the reverse. Since long-term retention is likely to be a function not only of the initiation-presentation (or presentation-initiation) interval but also of initial intensity of the process (which, for simplicity, is considered here as a step function), any exhaustive empirical determination is likely to be an onerous undertaking; efficiency would seem to be served if the attack were analytically based.

It should be possible to determine the separate contribution of enhanced discrimination and consolidation for the special, though common, case in which initiation and presentation occur simultaneously (e. g., speaker banging podium as he makes a critical point, learning material inducing the process). This determination would involve the successive application of presentation followed by initiation, with successive repetitions involving a reduction in the presentation-initiation interval. Under these conditions one would expect long-term retention to increase because the engram representation of the learning material would have had less of of an opportunity to deteriorate by the time the process begins to influence consolidation. By plotting long-term retention against presentation-initiation interval, and by extrapolating these values to zero interval, an estimate would be obtained of the amount of retention which would be expected from consolidation, independent of discrimination. If this amount is subtracted from the retention which results when presentation and initiation occur simultaneously, the difference should be a measure of the extent to which the process enhances discrimination at the time of initiation.

The procedure described above would require different groups of subjects at each presentation-initiation interval if the same learning material is used, or matched materials if the same group is used at each interval. And of course, each of these procedures requires a pre-adaptation procedure in order to eliminate resistance changes due solely to the presentation operation.

In the UM studies, presentation and initiation may be considered to have occurred at approximately the same time. Thus, it is not possible to determine the separate contributions of discrimination and consolidation to retention. However, this does not preclude a consideration of retention in relation to arousal. On this latter point, the UM results are quite clear: relative to low arousal, high arousal is associated with poorer immediate retention and better delayed retention. What about moderate arousal? Under moderate arousal, presumably, retention would fall between that associated with low and high arousal. Thus, the prediction would be that immediate retention associated with moderate arousal would be poorer than that associated with low arousal but better than that associated with high arousal. Similarly, delayed retention associated with moderate arousal would be better than that associated with low arousal but poorer than that associated with high arousal.

If in fact there is a process which results in not only a decrease in skin resistance but also an increase in discriminative ability and engram consolidation, then this process may be considered as enhancing learning, and would therefore have many of the characteristics attributed to "reinforcement," as this term is used in relation to external reinforcers and reinforcing conditions. Thus, a question is raised about the necessity of the concept of reinforcement when this concept is defined in terms of non-organismic factors. There would be no objection to defining reinforcement in terms of the process, or its possible concomitants, which might be expressed psychologically as "satisfaction," "pleasure," etc., or physiologically as parasympathetic compensation, amygdaloid stimulation (Olds, 1956), or other neural or biochemical activities (Berlyne, 1964), though this use of the term "reinforcement" would be tautological.

It is assumed here that the process subsumes reinforcement, allowing learning to be considered independent of the nature of the external reinforcer. If retention is influenced by by the process, then the nature of the initiator of the process is irrelevant, and learning should not be dependent on whether the process is initiated by the learning material, or by some other factor. This orientation is consistant with the position taken by Maltzman and Raskin (1965), who believe that the orienting response functions as a reinforcer. Their position is extended here to include a general process implied by all resistance decrements, not simply that decrement associated with the OR.

In summary of the preceding considerations, it is postulated that there is a process which, once activated, results in an increase in discriminative ability and engram consolidation. The onset of the process is manifested by a resistance decrement, and the result of the process is manifested by various measures of learning and retention. In particular, the (intensity of the) process is inversely related to immediate retention and directly related to delayed retention. Finally, the process constitutes "reinforcement," if this term is used, as it often is, to denote a process mediated by external reinforcers and reinforcing conditions.

With respect to the retention of information presented in an instructional communication, the preceding considerations seem to imply the following: initial accessibility and retentive duration of information are directly proportional to resistance decrement at or near the point in time at which the information was presented. Initial accessibility following information presentation is defined operationally as the first instant at which the subject can yield a behavioral indication of retention, usually one of recall or recognition. Retentive duration extends from that instant to a subsequent one at which the subject is no longer able to yield, at the chosen criterion level, a behavioral indication of retention.

In order to operationalize this formulation, it is convenient to consider two selected points in time following information presentation in order to define three types of events: forgetting events (measurable retention the first time, but not the second), retention events (measurable retention both times), and reminiscence events (measurable retention the second time, but not the first). There is, of course, a fourth type of event: no measurable retention at either time. This fourth type of event is not considered in the current study, since there would be no way of knowing whether the absence of a retentive measure is due to a failure in initial learning or due to the effect of arousal on retention.

The term "retention event" will be used in this report, even though it introduces still another usage of the word "retention," and hence, may not be the most apt. Hopefully the context will allow the reader to distinguish this usage of retention from the two other major meanings of the word: (1) as a process, similar to the process of remembering, and (2) as a concept, similar to the concept of memory.

Presumably there are two points in time following presentation-initiation which will yield all three types of events (forgetting, retention, and reminiscence). The UM studies suggest that an appropriate value for the first point is in the order of minutes, whereas the second point is in the order of days or weeks.

It is predicted that forgetting events are associated with small resistance decrements, retention events with moderate decrements, and reminiscence events with large decrements.

This prediction is tested in the current study. Two approaches are possible. First, resistance decrements at or near the time of information presentation would be assigned to three levels, and the expectation that small, medium, and large decrements would be associated with forgetting, retention, and reminiscence events, respectively, would be tested against observation. Second, forgetting, retention, and reminiscence events would be identified, and the expectation that these three events, respectively, would be associated with small, medium, and large resistance decrements at or near the time of information presentation would be tested against observation. The second approach is used in the current study.

In summary of this section, the theoretical implications are derived primarily from the important formulations of the UCLA studies and from the impressive UM studies involving forgetting and reminiscence events in relation to arousal. The current study, in addition, makes a prediction involving retention events in relation to resistance decrement, and considers retention in relation to presentation-initiation order.

1.32 Methodological Implications

Relative to the research reviewed in Section 1.1, the research reviewed in Section 1.2 had a greater influence in suggesting theoretical implications for the current study; the same is true with respect to methodological implications. The six studies in Section 1.1 differed methodologically from those indicated in Section 1.2 with respect to (1) the measurement of arousal, and (2) the specification of arousal level. These two problems will be discussed separately.

Arousal was measured verbally in the studies of Section 1. 1, physiologically in the studies of Section 1. 2. A comparison of Section 1. 1 and 1. 2 studies reveals that the latter type of measurement has two advantages.

The first advantage of physiological measurement of arousal is that the measure is obtained at the time of interest and without (or with minimal) interference with the experimental conditions. For human communication and learning studies in general, and most certainly for the studies of Section 1. 1, verbal measures of arousal can usually be obtained only by interrupting the normal presentation of material or by deferring measurement until the presentation reaches a stopping point, which is usually the completion of the presentation. Quite aside from dissimulation, responses at these points reflect the state of the subject at these points. This is true even though the subject is asked to respond as he felt at the time the material was being presented; his response reflects his memory at the time he responds. The UM results suggest that measures of feelings based on memory at the end of a communication are likely to be particularly in error for arousal-mediating communications.

The second advantage of physiological measurement of arousal is that the measure is more likely to allow comparison of results among studies. Since there is no standardization of verbal measurement of arousal, it is impossible to determine whether the same term, say high arousal, used by two experimenters refers to the same level of the phenomenon, assuming that they are referring to the same phenomenon. Recall that in each of the two studies in Section 1. 1 in which the result was significant and opposite to that predicted by the Janis-Feshbach hypothesis, the authors attempted to reconcile the discrepancy in terms of a "my high is really his low" explanation. Further, all but one of the studies in Section 1. 1 utilized content-specific measures of arousal. Under these conditions, comparison of results from studies involving different content becomes even more difficult. Content-independent measurement is more adequately achieved by current practices in physiological measurement of arousal. Standardization of physiological measurement at this time is probably premature, and in any case, it is usually possible to describe the measurement and analysis procedures with sufficient objectivity as to allow another investigator to duplicate these procedures in another content area (Levonian, 1962).

Even if there did not exist these two general advantages of physiological over verbal measurement of arousal, any verbal measure of arousal in the current study is likely to involve considerable error. The hypothesis of the study is such that the subject would have to recite at the end of the film his arousal change - not simply level - at each point in the film. In effect, the subject would be asked to report whether at, say, 2 minutes and 38 seconds relative to the beginning of the film, his arousal was increasing or decreasing as well as the time rate of this change. The studies in Section 1. 1 simply asked the subject to recall his arousal level over a broad portion of the communication; a test of the current hypothesis would require the subject to recall his arousal change at many points in the film. This latter task is considerably more difficult.

Because of the considerations of the preceding three paragraphs, the current study will utilize physiological measurement of arousal.

The studies in Section 1. 1 also differed from those in Section 1. 2 with respect to the specification of arousal. Each of the six studies in Section 1. 1 specified the two (three in the case of the Janis and Feshbach study) arousal groups in terms of the experimenter's designation of the arousal-mediating capacity of the two versions of the communication. Such a procedure does eventuate in the designation of one group as, say, the "high arousal group," but this designation pertains to the experimenter's perception, not to a characteristic of the group. Even then, not one of these studies obtained a measure of the reliability of this specification, or for that matter, even presented non-quantitative evidence indicative of reliability. Even stranger is the fact that in each of the four studies in which independently-obtained arousal data indicated a failure to reject the null hypothesis that the two groups were from the same arousal population, the experimenters continued to use the terms Strong (or High) and Minimal (or Low) arousal to identify the groups. Since these authors had taken the position that "my mind is made up; don't confuse me with the data," one wonders why they even bothered to obtain a measure of arousal.

In the studies of Section 1.2 the orientation with regard to the specification of arousal was simple; since arousal is a characteristic of the subject, the specification of arousal should be based on an arousal measure obtained from the subject. This orientation will be followed in the current study.

Had the studies of Section 1.1 decided to specify arousal on the basis of subject measures, there would have been no need to use two (three in the case of the Janis and Feshbach study) treatments. Each study apparently recognized a severe problem associated with the use of two treatments - namely, the possibility that an effectiveness difference between groups might be due to a difference between the treatments other than the difference in an experimenter-specified arousal-mediating capability. Thus, each study suggested that the two (or three) versions of the communication were similar except with respect to their arousal-mediating capabilities (e.g., Janis and Feshbach (1953), the versions "...differed only with respect to the amount of fear-arousing material presented"). Despite such claims, it is beyond the current state of the art to generate communications of the type used in Section 1.1 which differ only with respect to one variable. In any case, in view of the possible contamination which may be introduced by differential treatments they should be eschewed except when they are necessary (e.g., in a case in which multiple treatments are an integral part of the substantive hypothesis, in the randomization of an extraneous variable, etc.).

Each hypothesis in Section 1. 1 could have been tested with one experimental group; why this was not done remains a mystery, for not one of the studies justifies its use of more than one treatment. In any case, the current study will use only one treatment. Further, the main hypothesis in each Section 1. 1 study could have been tested without a control group; yet for reasons which are unclear to me, four of the six studies used a control group. In any case, since a control group is not essential to a test of the hypothesis in the current study, a control group will not be used.

In summary of this section, the methodology employed in the current study is based on the studies of Section 1. 2, despite the fact that the current study falls into the class of studies in Section 1. 1 - that is, a study dealing with communication effectiveness in relation to arousal.

2 Method

2.1 Subjects

Subjects consisted of driver education students enrolled in a high school in Los Angeles. There were four classes, and each met in the same room. The classes met about one hour apart, with two meeting before lunch, the other two after lunch. Practically all of the students were in the tenth grade and were 15 years old. The study is based on the 83 subjects for whom complete data were available.

2.2 Procedure

The skin resistance of each student was recorded as he watched a traffic safety film entitled Safety Belt for Susie. This film had been scheduled for showing in class as an integral part of the course. Data collection conformed to this schedule. Data were collected in the room and at the time the classes normally met, and students sat at the seats assigned to them for the entire semester.

After the teacher made some preliminary announcements, he introduced the experimenter, who introduced the study as one designed to measure student's physiological reactions to traffic safety films. The experimenter demonstrated the manner of sliding a pair of finger-tip electrodes (Levonian, 1963b) onto the index and ring fingers of the non-writing hand, and asked the students to do likewise with the "sensor" (the term "electrode" was not used) being handed out by an assistant. Proper positioning and pressure were clarified by the experimenter, and the student was encouraged to ask for another-sized ring if the first wasn't entirely comfortable.

After the rings were in place, ten minutes were allowed to intervene before the projector was turned on and data recorded, so as to allow for stabilization of the perspiration under the electrodes. During this time the student was asked to attach the cable from his sensor to the cable terminating at his desk. In order to fill in the 10 minutes, irrelevant matters were discussed, such as the manner in which the number on the student's cable allows his skin resistance record to be identified, the material from which the sensor is fabricated, the fact that man doesn't have sensory mechanisms to detect resistance changes directly, etc. The lights were turned out and the projector started. The students were told to relax, since it would be a full minute before anything appeared on the screen. This was intended as a period during which the student would adapt to the sound of the projector.

Immediately after the 10-minute film, a 30-item questionnaire was passed out. The items pertained to the film, and the 15 items which pertained to information presented at specific points in the film appeared in the questionnaire in essentially the same order as they appeared in the film. Each of these 15 items was answered about 10 minutes after presentation of the information pertaining to the item - that is, it took about 10 minutes to fill out the questionnaire. The students were thanked for their cooperation, and asked to leave their names and addresses if they wished a non-technical summary of the results mailed to them. The impression was left that the data collection phase was completed.

A week later the experimenter appeared before the class, and requested that the same questionnaire be filled out again as a measure of retention. The questionnaire was identical in form, and it was introduced as such.

2.3 Questionnaire

The 15 items pertaining to specific points in the film are given in Table 2.1. The item numbers pertain to the original questionnaire.

Each item had two choices - yes and no - and the student was asked to guess if he didn't know. The correct answer is given in parenthesis after the item.

TABLE 2.1

QUESTIONNAIRE

4. Nancy's parents were involved in an accident in which their car hit a tree. (yes; visual; 4360)

5. Nancy was not with her parents during their accident because she was in school. (no; auditory; 3640)

6. At the time of the accident there were no other cars on the highway. (no; visual; 4200)

7. The accident resulted in a broken windshield. (yes; visual; 4320)

9. Nancy's parents had the name of Norwood. (yes; auditory; 5320, 5840)

11. The doctor took Nancy's parents to see the experimental collisions. (no; visual; 6400)

15. All "passengers" in the experimental collisions wore safety belts. (no; visual; 7200, 7880)

16. At least one of the cars in the experimental collisions had "UCLA" painted on its side. (yes; visual; 8360, 8520, 9120)

17. The film stated that children should never be allowed to stand on the seat of a car. (yes; auditory; 9280)

18. In many of the experimental collisions where the children were not restrained, medical diagnosis showed that they suffered only minor injuries. (no; auditory; 10080)

19. The film indicated that an adult safety belt is not effective for children below the age of five. (no; auditory; 10520)

22. One of the dolls in the experimental collisions was named Carla. (yes; visual; 11600, 12880)

23. The film stated that a driver should admit to himself that an accident could happen to him. (yes; auditory; 11640)

24. At least one car in the experimental collisions had its lights on before impact. (yes; visual; 8280, 12440)

27. The film was directed toward driver education students. (no; auditory; 14680)

The 15 items consisted of 7 auditory items (items pertaining to information contained only in the sound track and 8 visual items (items pertaining to information contained only in the picture). Item type is indicated in parenthesis after the item.

Also in parenthesis is the frame number at the midpoint of the presentation of the information, with this critical frame rounded to the closest multiple of 20. For each item, the information was presented within ±50 frames of the critical (midpoint) frame. For some items the same information was presented at more than one point in the film, and the critical frame numbers associated with each of these points are indicated in Table 2.1.

The remaining 15 items in the questionnaire pertained to information presented over an interval of more than 100 frames. These tended to be general items such as, "The main theme of the film was to drive carefully."

2.4 Film

Safety Belt for Susie[1] was produced by Charles Cahill and Associates, Inc., in 1962. The film is based on collision injury research conducted by Derwyn Severy of the Institute of Transportation and Traffic Engineering, University of California, Los Angeles, and sponsored by the U.S. Public Health Service.

The film consists of three parts. The first part lasts about 4 minutes, and in addition to head titles, establishes an attractive family consisting of a 4-year old girl and her parents, and emphasizes the similarity between the daughter and her doll, Susie: they are roughly the same size, wear identical clothes, ride together on the ferris wheel, etc. With the daughter elsewhere, the parents are involved in a collision. Although they are only slightly hurt, Susie is thrown over the front seat; the film indicates that had the daughter been there, she too would have been propelled and would have sustained serious injuries. The parents are examined by their physician, who proceeds to tell them of his participation in the UCLA collision injury research. This fades into the second part of the film, which lasts about 5 minutes and revolves about activities focussed on experimental collision injury research. The film shows a number of full-scale collisions involving anthropometric dummies, some children. The third part of the film, lasting no more than 2 minutes, returns to the original family, showing its members, including Susie, utilizing their newly-obtained safety belts. The 15 questions were taken from each of the three parts.

---

[1] This film is available for rental or sale from the University of California Extension Media Center, 2223 Fulton St., Berkeley, California 94720.

Frame 1 preceded the first frame of the head fade-in by 240 frames; thus, the first frame of the film proper was Frame 241. The last frame of the tail fade-out was Frame 15840, but skin resistance data were obtained for an additional 240 frames - through Frame 16080.

2.5 Skin Resistance

The technique for recording skin resistance has been described elsewhere (Levonian, 1962), and only a brief summary will be given here.

Each student wore a pair of stainless steel electrodes on the distal phalanges of the index and ring fingers of his non-writing hand. Each electrode was 3/8 inch square, and curved to conform to the volar surface of the finger. A constant current of 2 microamps was passed through the skin, and the resulting DC voltage was sampled, amplified, and recorded on magnetic tape. The signal from each student was sampled each time a film frame appeared on the screen - roughly 24 times a second - and the sampled value recorded in analog form on a separate channel. Later the recorded values were converted to digital (11-bit straight binary) and transmitted by direct wire to a computer, allowing the data to be recorded on magnetic tape in computer format. The conversion time was 100 microseconds.

Since the film contained 16080 frames, there were 16080 resistance values for each subject. The data for each student were standardized to a mean of zero and a variance of unity.

3 Results

It should be recalled that all tests of significance in this report are referred to the .05 level, two tailed, with the exception that in the current study, all tests of significance are referred to the .01 level, two tailed.

The prediction of this study is that forgetting, retention, and reminiscence events will be associated with, respectively, smaller, medium, and larger resistance decrements. In the test of this prediction it is necessary to define operationally the terms "event" and "decrement".

The terms "forgetting event," "retention event," and "reminiscence event" are each defined in a similar manner. Consider a student who shows forgetting of the information asked in a particular item (correct response at the 10-minute retention interval, incorrect response at the 1-week retention interval). The forgetting of this information is defined as a forgetting event. A student may have more than one forgetting item, in which case, all such items are considered as a single forgetting event; thus, a student may be associated with at most one forgetting event. Retention and reminiscence events are similarly based on retention items (correct responses at both retention intervals) and reminiscence items (incorrect response at the 10-minute retention interval, correct response at the 1-week retention interval). So measured, reminiscence in this report pertains to the Ballard-Williams phenomenon, not the Ward-Hovland phenomenon (Osgood, 1953).

The term "decrement" is defined operationally relative to an event, as defined in the previous paragraph. Consider a student who has only one forgetting item. Associated with that item is a critical film frame (see Section 2.3). The student's resistance decrement is the maximum decrease in his skin resistance over a decrement interval whose midpoint is the instant at which the critical frame appeared on the screen. The student's decrement for this interval is taken as the maximum difference between any resistance value in the interval and any subsequent value which is lower and in the interval. If the student shows no resistance decrease in the interval, the "decrement" is taken as zero. If the student's forgetting item is associated with more than one critical frame (more than one place in the film), the largest decrement is taken as the student's forgetting decrement. If the student has more than one forgetting item, with a decrement for each item, the average of these decrements is taken as the student's forgetting decrement. Retention and reminiscence decrements are similarly defined.

The preceding two paragraphs make it clear that each forgetting event has associated with it one and only one decrement (which may be referred to as a forgetting decrement), and that a student yields at most only one forgetting event and its associated decrement. Similar statementa can be made with respect to retention and reminiscence.

In this study there were 42 forgetting events (42 students showed forgetting of at least one item), 83 retention events, and 61 reminiscence events. Of the 83 students, 30 showed all three types of events.

3.1 Type of Event in Relation to Decrement

It was hypothesized that forgetting, retention, and reminiscence events would be associated with, respectively, smaller, medium, and larger resistance decrements at or near the time of information presentation.

In the test of this hypothesis, forgetting, retention, and reminiscence decrements were determined for each of the 30 students who showed all three types of events. These three values were determined for a decrement interval of ±10 seconds, and then determined again for decrement intervals of ±20, ±30, ±40, ±50, and ±60 seconds.

If there were no relation between decrement and type of event, one would expect only 5 of the 30 students to show the particular permutation predicted - that is, increasing decrements for forgetting, retention, and reminiscence event, respectively. The actual number of such students is given in Table 3. 1.

TABLE 3. 1

NUMBER OF STUDENTS SHOWING INCREASING DECREMENTS FOR FORGETTING, RETENTION, AND REMINISCENCE EVENTS, RESPECTIVELY

Decrement Interval	Number of Students	Chi Square (Yates' Corrected)
±10	7	. 54
±20	8	1. 50
±30	11	7. 26*
±40	11	7. 26*
±50	10	4. 86
±60	10	4. 86

*significant

The results shown in Table 3. 1 are readily summarized. (1) For each decrement interval, the results are in the expected direction. (2) As the decrement interval increases to ±30 and ±40 seconds, chi square increases, then decreases. (3) Chi square is significant for the ±30 and ±40 second decrement intervals.

The results shown in Table 3. 1 are taken as supporting the hypothesis that there is a process which affects skin resistance and event type. The results indicate that the relationship between event type and resistance decrement is sufficiently reliable as to allow the prediction of the latter from a knowledge of the former. Presumably the converse prediction could have been made also, though no attempt to do so was made in the current study.

A determination was made also of the resistance level (not decrements) associated with each of the three types of events. The results would not allow a rejection of the null hypothesis that the resistance levels associated with forgetting, retention, and reminiscence events were from a single population of resistance levels. A similar interpretation resulted from an analysis involving resistance increment. These intra-subject analyses, with the analyses of resistance decrement, were based on the standardized data.

Are these results explicable in terms of a coincidental group occurrence of event type and resistance decrement? If it were true that items associated most frequently with forgetting, retention, and reminiscence events were associated, respectively, with smaller, medium, and larger resistance decrements for the group as a whole, then the results of this study would be less clearly interpretable in terms of an individual process which affects both skin resistance and learning-remembering.

Consider, for instance, the possibility that many subjects forgot information which just happened to be presented at a time when most subjects yielded smaller resistance decrements. However, it if were actually true that the forgetting of this information was only coincidentally related to smaller decrement, then one would expect the few subjects who showed reminiscence for this information to also yield a smaller resistance decrement. Thus, for a given item, one would expect to find no sytematic difference in decrement between students who show forgetting and reminiscence.

However, a comparison of forgetting and reminiscence students for each item does reveal a systematic difference between these two groups with respect to resistance decrements associated with each item. These results are given in Table 3. 2. Each entry is a z-score. A particular entry represents a decrement averaged over the students in that category, with the number of students in each category given in Table 3. 3. For instance, for Item 6 there were two forgetting students, and the average

TABLE 3.2

RESISTANCE DECREMENT AVERAGED OVER STUDENTS SHOWING FORGETTING AND REMINISCENCE FOR EACH ITEM

Item	± 10		± 20		± 30		± 40		± 50		± 60	
	For.	Rem.	For.	Rem.	For.	Rem.	For.	Rem.	For.	Rem.	For.	Rem.
4												
5												
6	.29	.38	.29	.46	.29	.52	.29	.69	.29	.94	.30	.99
7	.53	.93	.64	1.12	.72	1.15	.86	1.41	.86	1.62	1.02	1.67
9	.28	.61	.43	.77	.54	1.04	.74	1.12	1.16	1.42	1.30	1.46
11												
15												
16	.54	.54	.64	.75	.64	.84	1.48	.89	1.52	.90	1.53	.91
17	.36	.15	.86	.15	1.11	.15	1.14	.46	1.18	.87	1.44	1.95
18	.18	.46	.26	.46	.31	.46	.37	.48	.40	.48	.52	.48
19	.50	.30	.62	.35	.79	.44	.93	.57	.98	.66	1.02	.68
22	.67	.74	1.56	1.69	1.56	1.90	1.82	2.00	2.42	2.23	2.49	2.36
23	.14	.46	.14	.63	.15	.78	.70	1.20	.70	1.49	.78	1.77
24	.10	.75	.10	1.11	.42	1.48	.61	1.77	.67	1.83	1.13	1.88
27	.41	.47	.57	.71	.89	.95	1.06	1.19	1.10	1.35	1.52	1.53
Weighted Average	.41	.58	.57	.82	.72	1.02	.90	1.21	1.00	1.35	1.23	1.45

TABLE 3.3

SAMPLE RESISTANCE DECREMENT IN RELATION TO THE NUMBER OF STUDENTS SHOWING FORGETTING, RETENTION, AND REMINISCENCE FOR EACH ITEM

	Sample Resistance Decrement						Number of Students		
Item	±10	±20	±30	±40	±50	±60	For.	Ret.	Rem.
4	.29	.29	.36	.36	.36	.50	0	77	2
5	.03	.23	.40	.50	.50	.50	1	78	0
6	.29	.36	.36	.36	.50	.50	2	77	1
7	.29	.36	.36	.36	.45	.50	2	74	3
9	.07	.07	.10	.11	.29	.29	6	37	18
11	.08	.10	.11	.14	.14	.14	5	60	0
15	.18	.26	.32	.32	.32	.32	1	80	0
16	.14	.32	.32	.32	.32	.32	2	45	21
17	.03	.08	.08	.08	.14	.32	4	76	2
18	.01	.09	.09	.09	.09	.09	3	77	2
19	.09	.09	.09	.09	.10	.23	21	30	4
22	.23	.30	.35	.35	.35	.42	1	79	2
23	.23	.23	.23	.28	.31	.35	2	67	9
24	.20	.32	.32	.41	.46	.46	1	15	22
27	.15	.15	.24	.32	.41	.53	5	43	10
							56	915	96

of their decrements is .29 for the ±10-second retention interval. In Table 3.2 entries for Items 4, 5, 11, 15 have been omitted because of the absence of forgetting or reminiscence for those items. In computing the weighted average shown in the last row, each of the 11 column entries was weighted by the number of students contributing to that entry.

It should be noted that the results in Table 3.2 are not based solely on the 30 students of Table 3.1. Any student who showed forgetting without reminiscence, or vice versa, is included in Table 3.2, and this includes fully 90% of the students.

In general, the results in Table 3.2 reveal that students who show forgetting for an item tend to have a smaller resistance decrement than do students who show reminiscence for the same item. Since both decrements occurred simultaneously, it seems unlikely that the results in Table 3.1 are due to any obvious group-related artifact.

This same conclusion is reached by comparing separately for each item (1) the sample resistance decrement associated with the item with (2) the number of students showing forgetting of the item, retention of the item, and then reminiscence of the item. This information is given in Table 3.3. Sample resistance decrement represents an average over the sample, and is obtained by summing the individual standardized records, then dividing by 83 - that is, by summing the 83 ipsatively-standardized values associated with each of the 16,080 film frames, then dividing each of these 16,080 sums by 83.

Table 3.3 does not show any obvious relation between resistance decrements for the sample as a whole and the number of students showing forgetting, retention, and reminiscence of each item. Thus, it would appear that the intra-subject relation between resistance decrement and event type is not explicable in terms of a sample trend.

Table 3. 3 also suggests that the interaction between resistance decrement and event type is not explicable in terms of an order effect. There appears to be no obvious relation between the number of students of each event type and the order of the item, which is also the same as the order in which the item information was presented in the film.

It should be noted that the sample decrements in Table 3. 3 tend to be smaller than the decrements shown in Table 3. 2. The decrements shown in Table 3. 3 are averaged over 83 subjects, whereas those in Table 3. 2 are averaged over fewer subjects, as given in the Forgetting and Reminiscence columns of Table 3. 3. This general tendency for decrements to shrink when based on a larger number of subjects is a reflection of the variability among subjects with respect to the temporal trends of their resistance records. There was very little in the way of a resistance pattern common to all, or even most subjects; the resistance of one subject might be increasing at the same time that the resistance of another might be decreasing. There appeared to be no strong general tendency. For instance, while there appeared to be a tendency in some cases for a resistance decrement to be greater if it followed a rising resistance, a phenomenon previously noted by Cattell (1928), there were many exceptions.

The general finding was one of wide variability among subjects. This variability tends to eliminate a sample trend. Statistical evidence for this fact is revealed by the variance in the sample resistance record, consisting of 16, 080 values, each averaged over 83 subjects. The variance of this sample record is . 025, whereas the variance of each of the 83 contributing records is 1. 000. Of course, the mean of the sample record is zero, since this is the value of the mean of each contributing record.

Before leaving Table 3. 3, it should be noticed that this table allows the determination of the number of students who failed to yield the correct answer at either of the two retention intervals; this number is obtained by subtracting from 83 the sum of the other three types of students associated with the item. For instance, 4 students failed to give the correct answer to Item 4 at either retention interval.

3.2 Temporal Relation of Decrement to Presentation

Section 1.31 included a consideration of retention as a function of the temporal relation of initiation to presentation. It is recalled that "initiation" pertains to the initiation of a change in skin resistance, while "presentation" pertains to the presentation of information. It was argued that there is currently no way of determining analytically the effect of initiation-presentation order on retention.

However, a preliminary investigation of this problem area can be made using the data of this study. Consideration will be given to the following question: Are forgetting, retention, and reminiscence events influenced by decrement location? The term "decrement location" pertains to the temporal occurrence of the decrement relative to the presentation.

The word "decrement" is used instead of "initiation", for the current study utilizes a skin resistance measure based on a resistance change which might occur anywhere within a broad but specific interval; the word "initiation" is more appropriate for the case in which the measure is based on a change whose beginning is referred temporally to a specific stimulus event. Since the term "GSR" is often applied to some variant of this latter measure, it should be apparent that the resistance measure used in this study is not a measure of GSR. That is, resistance decrement in the current study is not necessarily considered as a response.

It is recalled that the decrement interval covers equal durations (say, ±10 seconds) before and after the instant at which the critical frame appears on the screen. In this section, this decrement interval will be divided in half, allowing the specification of a minus-interval (-10 seconds) computed from the beginning of the decrement interval to, and including, the critical frame, and a plus-interval (+10 seconds) computed from the frame immediately following the critical frame through the end of the decrement interval. The corresponding decrements will be referred to as minus-interval decrement and plus-interval decrement.

The data analysis in this section was quite simple. For each of the 30 students who showed each of the three events, a determination was made of his 10-second minus-interval forgetting decrement and for his 10-second plus-interval forgetting decrement. The latter decrement was subtracted from the former, yielding 30 differences. Then a test was made of the null hypothesis that these 30 differences are from a population of differences whose mean is zero. Finally, this procedure was repeated for minus- and plus-intervals of 20, 30, 40, 50, and 60 seconds, and for retention and reminiscence decrements.

The results are given in Table 3.4. A negative sign indicates that the mean of the 30 minus-interval decrements was smaller. It is seen that in no case does minus-interval decrement differ significantly from plus-interval decrement.

In Table 3.4 there is a tendency for (a) retention events to be associated with larger minus-interval decrements, and for (b) reminiscence events to be associated with larger plus-interval decrements. With respect to decrement, the interaction between event type (retention vs. reminiscence) and interval (minus vs. plus) is significant, in general.

This result is consistent with the consideration in Section 1.31 pertaining to decrement initiation-information presentation order. It was suggested there that a resistance decrement which occurs before information presentation reflects a process which enhances both learning (via enhanced discriminative ability) and remembering (via enhanced engram consolidation), whereas only the latter is enhanced by a resistance decrement which follows information presentation.

TABLE 3.4

DIFFERENCE BETWEEN MINUS-INTERVAL AND PLUS-INTERVAL DECREMENTS

(N=30)

Decrement Interval	Event Type	Mean of Decrement Difference	SE	t
10	For.	1.80	4.86	.37
	Ret.	3.90	2.23	1.75
	Rem.	-12.13	7.10	1.71
20	For.	7.80	5.92	1.32
	Ret.	6.07	2.69	2.25
	Rem.	-17.47	9.30	1.88
30	For.	-2.47	8.17	.30
	Ret.	4.87	3.84	1.27
	Rem.	-17.60	8.94	1.97
40	For.	-4.13	9.26	.45
	Ret.	5.63	4.01	1.40
	Rem.	-17.87	11.35	1.57
50	For.	-9.43	8.35	1.13
	Ret.	8.33	4.20	1.98
	Rem.	-14.77	13.05	1.13
60	For.	-10.73	8.91	1.20
	Ret.	14.03	6.41	2.19
	Rem.	-18.83	14.60	1.29

The compatibility of this suggestion with the results of Table 3.4 may be seen by considering a resistance decrement which precedes information presentation. Such a decrement is assumed to enhance both learning and remembering, with the former reflected mainly in short-term retention, and the latter mainly in long-term retention. Now consider a decrement which succeeds presentation. Such a decrement is assumed to enhance only remembering, a process which is assumed to take time, resulting in long-term retention only.

4 Discussion

Of the following sections, the first introduces the concept of information accessibility. Information accessibility considers information retention as a continuum, while at the same time allowing for the distinction between retained information which is, and is not, manifested behaviorally, say by recognition or recall. Information accessibility emerges as the integrative concept of a model intended to explain the results of this study.

The next section first discusses two classes of information storage theories, and then presents a biological explanation which is consistent with the information accessibility model.

The results are then compared with the two types of studies presented in Section 1 in order to provide a broader base for the interpretation of the results which emerged in those studies and in the current study.

Finally, a presentation is made of the general implications of the results, as well as the more specific implications to traffic safety films.

4.1 Information Accessibility

Information accessibility is defined here as the extent to which information which is stored, or being stored, in the subject is available to the subject. In general, information accessibility would be measured behaviorally, and most typically by recall or recognition, as in this study. It is assumed that information accessibility has a behavioral threshold. A subject is above this threshold if he is able to yield a behavioral indication of the test information.

It is assumed here that information accessibility is a joint function of (1) a process, one of whose manifestations is skin resistance change during information presentation, and (2) time after information presentation. The form of this information accessibility function might be something like that shown in Figure 4. 1. It is seen that the information accessibility surface is a joint function of (1) resistance decrement, the reduction in skin resistance over a short interval spanning the point in time at which information is presented, and (2) retention interval, the time between the presentation of information and the test of retention.

It is postulated that forgetting, retention, and reminiscence events may be discussed meaningfully in terms of Figure 4. 1.

Three lines, coming toward the viewer, have been drawn on the information accessibility surface. The middle line connects the maxima of the family of information accessibility-retention interval curves, while the two outer lines connect the behavioral threshold points, points above which information accessibility can be measured behaviorally. It is seen that each information accessibility-retention curve is intersected by the behavioral threshold plane at two points.

Figure 4. 1 is hypothetical, of course, but it may be fruitful to attempt to interpret the results of this study in terms of this figure. First, certain characteristics of the manifold should be noticed. (1) As resistance decrement increases, maximum information accessibility increases. (2) As resistance decrement increases, maximum information accessibility occurs at a later point in time. (3) As resistance decrement increases, the two behavioral threshold points for a given information accessibility-retention interval curve occur at later points in time; the rate of this increase in time is greater for the threshold point on the falling part of the curve than it is for the point on the rising part of the curve. And (4), for vanishingly small values of resistance decrement (essentially indicative of the absence of the process), information accessibility fails to reach the behavioral threshold for any retention interval.

The model shown in Figure 4. 1 is consistent with the results of this study. This is more readily seen if it is noted that the three information

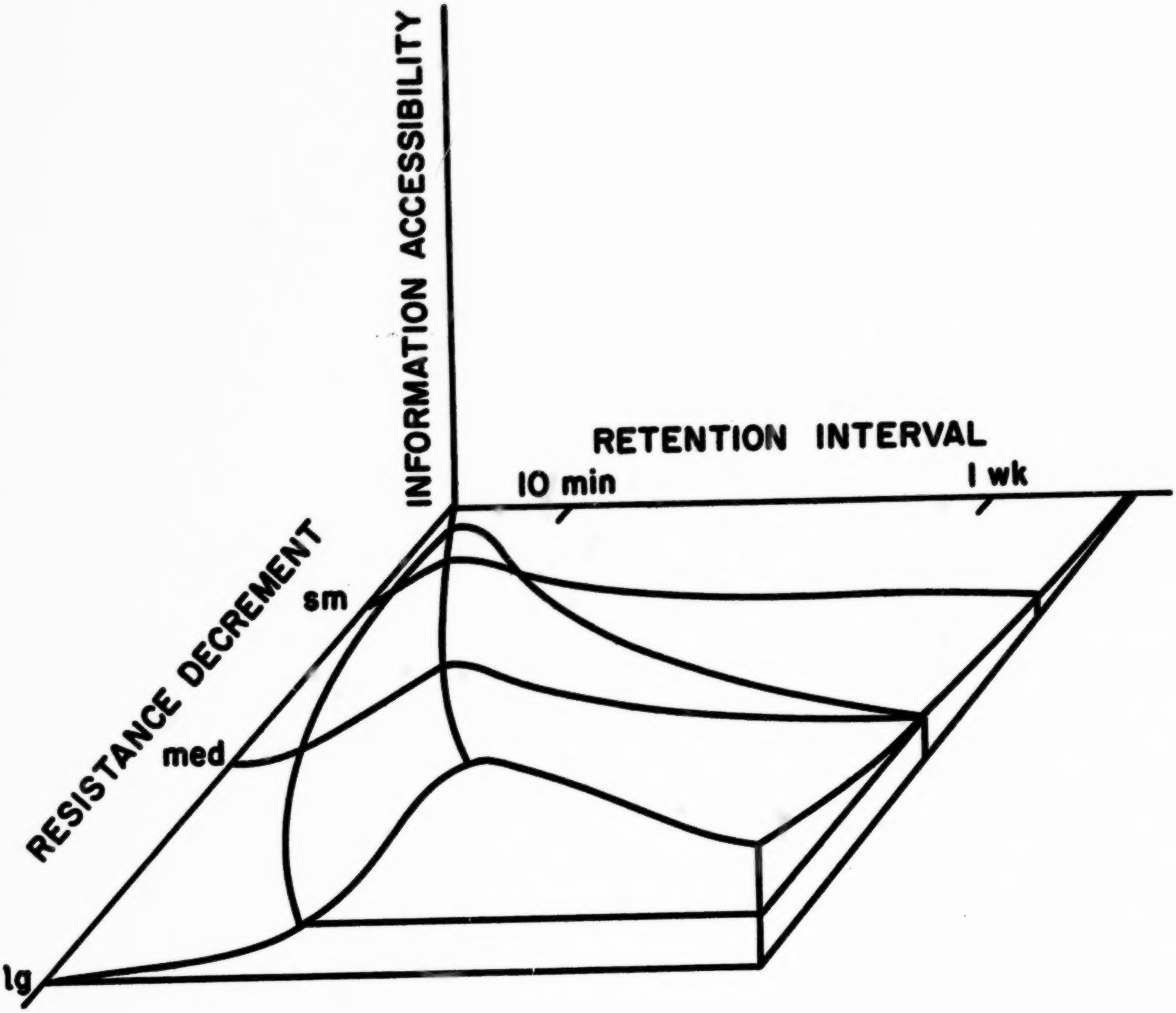

FIGURE 4. 1

INFORMATION ACCESSIBILITY AS A FUNCTION OF RESISTANCE DECREMENT AND RETENTION INTERVAL accessibility-retention interval curves in Figure 4. 1, corresponding to the three levels of resistance decrement, are associated with the three types of events defined in this study. The curve corresponding to the smaller resistance decrement is associated with forgetting events, the medium resistance decrement with retention events, and the largest resistance decrement with reminiscence events. At the end of the 10-minute retention interval, the forgetting and retention curves are above the behavioral threshold, whereas the reminiscence curve is not. At the end of the 1-week retention interval, the forgetting curve has dropped below the threshold, the retention curve remains above the threshold, and the reminiscence curve has risen will above the threshold.

The model shown in Figure 4. 1, together with the results of this study, suggest the generalization that a larger resistance decrement during exposure to material allows that material to be accessible (behaviorally available) over a longer period of time.

The fruitfulness of any model must be based on the studies it suggests, the success with which it predicts the results of these studies, and the number of points at which the model can be attacked.

The model suggests a study which would involve three measures of retention - say the two used in the current study and a third at a retention interval of one month. The 10-minute and 1-week retention measures would be used to identify forgetting, retention, and reminiscence items, as in the current study. The model would predict that only the reminiscence items would be above the behavioral threshold at one month.

The model suggests another study which would also involve three retention intervals -- say at one minute, one hour, and one week. Using the first two retention intervals to identify retention items (information accessibility above the behavioral threshold at both intervals), and considering two such items, one associated with a smaller resistance decrement and the other with a larger, the model would predict that only the larger decrement item would be above the behavioral threshold at one week. A similar prediction may be made for two reminiscence items.

These two suggested studies have involved three retention intervals, the first two being used to define type of item, and the third retention interval being used to test some prediction regarding the accessibility of these items at that third point in time. Other studies might utilize many more retention intervals.

Consider, for instance, a study intended to plot the information accessibility surface. Its height, standardized to unity, may be taken as the proportion of correct answers. Items would be assigned to categories, say three, according to the subject's resistance decrement at the time of information presentation. By repeating the retention test at a number of retention intervals, the information accessibility surface would be revealed. The result would be ipsative if only a single subject were involved, normative if a sample were involved. If a sample is used, initial ipsative standardization of resistance decrement, as in the current study, is necessary before items are categorized.

With certain types of information and items, one might suspect that repetition of the retention test would introduce a contaminating practice effect. There are at least two methods for precluding this possibility. The first method requires the use of a different sample for each retention interval. The second method requires a large number of items in each resistance decrement category. Then, the items within each category would be randomly assigned to n groups, one for each of the n retention intervals. While the first method is restricted to a normative surface, the second method is appropriate to either an ipsative or a normative surface. Again, however, ipsative standardization of decrement is necessary if a normative accessibility surface is desired.

Once the information accessibility surface is generated, multiple presentation of the retention test can be eliminated. Thereafter, knowledge of resistance decrement near the time of information presentation allows a prediction of whether accessibility is above the behavioral threshold for any particular retention interval. If a sample of subjects is involved, the proportions of correct responses associated with several categories of resistance decrements may be taken as the relative heights of the surface at the particular retention interval employed.

In view of the variation in resistance measuring techniques, articulation among studies would be enhanced if the practice were followed of standardizing each subject's resistance record to the same mean and variance. Similarly, the information accessibility surface would be standardized if the practice were followed of specifying information accessibility in terms of the proportion of test information above the behavioral threshold.

4.2 Biological Considerations

For simplification, the relation between resistance decrement and retention has been attributed in this study to a "process." A consideration of the possible nature of this process is more appropriately undertaken within a broader consideration of the biological aspects of retention.

There are two fundamental questions. First, how is information stored? And second, how is information retrieved? These questions would probably be phrased differently by persons who are biologically, physically, or psychologically oriented, but all would agree that the answers are not known.

A great deal of effort is directed currently toward the first question, biologically expressed: How is sensory information recorded (encoded) biologically? Much less effort is directed currently toward the second question, and it is not at all clear at this time that it is a question which can be answered biologically. Indications are propitious for a biological explanation of information storage, but the attempts to explain information retrieval may eventuate in a new orientation which is neither biological nor psychological. Because of the considerations in this paragraph, the discussion in this section will be focussed on possible biological bases of the storage aspect of retention; the retrieval aspect will be mentioned only, at the end of the section.

There are several classes of biological theories of information storage. Concentration here will be on molecular theories and on cytological theories. Some of the molecular theories are analogs of modern genetic theory; cytological theories have a longer history, are more physiologically oriented, and are amenable to the incorporation of mechanisms compatible with modern genetic theory (chemical genetics). Recent excellent reviews of these theories are available (Gaito and Zavala, 1964; Pribram, 1960).

Interest in the current report is restricted to arousal in relation to these two classes of theories. This topic is discussed more readily after a brief review of each class of theories.

Molecular theories assume that the memory engram is represented by the formation or modification of some type of molecule. The molecules most commonly suggested are deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins, and lipids. Since neuronal DNA and RNA have received the greatest attention, only these two will be considered here.

These genetically-oriented molecular theories suggest that the mechanism by which DNA provides a genetic code via the linear sequence of bases may have an analog for memory (Gaito, 1961; Gaito, 1963; Hyden, 1961). An evaluation of these theories must be based on genetic mechanisms, which are summarized in the next three paragraphs.

All cells, including neurons, contain DNA and RNA, with at least one type of DNA molecule for each gene. Each DNA molecule consists of a long, unbranched chain of nucleotides. In somatic cells, each DNA molecule is chemically bonded with a complementary molecule, and the two molecules form a twin helix. DNA replicate only upon cell division, at which time each component of the helix acts as a template for the synthesis of a complementary polynucleotide. DNA are confined to the nucleus of the cell.

Nuclear DNA act as templates for the synthesis of messenger RNA, which then migrate into the cytoplasm. There, messenger RNA are incorporated in ribosomes, which may be considered as protein factories. Transfer (soluble) RNA transport amino acids to the ribosomes, after which the transfer RNA are released intact to forage for more amino acids; each of some 20 types of transfer RNA recognizes its particular amino acid. Messenger RNA act as templates for the synthesizing of proteins from these amino acids. Similarly, viral RNA act as templates for the synthesis of enzyme proteins. Hence, the sequence of amino acids in a particular protein is determined indirectly by the base sequence of the parent DNA.

Since each somatic cell contains the same genes, and since each gene controls the synthesis of a particular enzyme (or other protein) by the mechanism described above, each somatic cell has the potentiality of synthesizing all enzymes. However, enzyme concentrations vary widely in different tissues, but the mechanisms which control these differences are obscure at present.

With this brief overview, we are ready to consider the particular molecular theories of interest here - namely, those based on DNA and RNA formation or modification. These theories assume that memory is represented by induced or modified neuronal DNA and RNA, particularly the latter.

The most direct experimental evidence which supports RNA induction or modification indicates that the base ratio of nuclear RNA in rats is changed with a specific learning task, that this change persists for at least two days, and that the change is associated with both neuronal and associated glial cells (Egyhazi and Hyden, 1961). This evidence will be considered first in terms of induction, then in terms of modification.

While the evidence that learning is associated with a change in base ratio is compatible with the interpretation of polynucleotide formation, such an interpretation is incompatible with current genetic theory. All nucleic acids, DNA as well as all known types of RNA (messenger, viral, transfer, and ribosomal) are formed on cellular templates according to complementary nucleotide sequences. It is known that no polynucleotide synthesis occurs in the absence of a template (Kornberg, 1961), and this evidence would seem to preclude any explanation of the altered RNA base ratio in terms of nucleic acid formation.

The altered base ratio result is compatible also with a polynucleotide modification explanation based on mutation of the RNA or the parent DNA. While mutagenic agents can alter chemically the purines and pyrimidines of the polynucleotide, the mutation is one in which some of the bases are converted to other related compounds not normally found in polynucleotides. However, this explanation appears to be precluded by the fact that no unusual bases have been found in brain DNA or RNA (Briggs and Kitto, 1962).

Another type of polynucleotide modification has been suggested - a rearrangement of nucleotides. While it is conceivable that electrical impulses might rearrange the order of the sequence, there is no direct experimental evidence for this.

Because of these and other considerations, genetically-oriented molecular theories of memory based on polynucleotide formation or modification have been rather strongly criticized (Briggs and Kitto, 1962; Dingman and Sporn, 1964; Goldberg, 1964). This criticism is directed at the notion of a "memory molecule" represented by an induced or modified RNA molecule, and not against the notion that RNA is involved in learning and memory.

In fact, there is considerable evidence to indicate that RNA is involved in learning-memory. First, ribonuclease blocks the retention of a conditioned response in regenerated planarian tails (Corning and John, 1961). Without this treatment, a planarian cut in half after learning a response, will show partial retention of the response for both regenerated halves. Second, administration of a substance similar structurally to guanine, a purine base, affects maze learning in the rat. The treatment does not affect retention of previous maze learning, but it does impair subsequent maze learning (Dingman and Sporn, 1961). Third, long-term treatment with yeast RNA facilitates operant learning in the rat (Cook, Davidson, Dixon, Green, and Fellows, 1963). And fourth, yeast RNA administered intravenously or orally over a long period of time improves the memory performance of older humans with a memory disturbance having an arteriosclerotic basis (Cameron, 1963; Cameron, Sved, Solyom, Wainrib, and Barik, 1963).

Attempts have been made to interpret the above results in terms of established cytological functions - this because of the difficulties believed to be associated with genetically-oriented molecular theories of memory (Briggs and Kitto, 1962) or, for that matter, all molecular theories of memory (Dingman and Sporn, 1964). It is held that the evidence adduced in favor of memory molecule theories is consistent also with more traditional cytological theories, and that, therefore, this evidence is insufficient to distinguish between the two classes of theories (Briggs and Kitto, 1962; Dingman and Sporn, 1964). For instance, the fact that learning is associated with a change in base ratio does not necessarily imply that learning is associated with modified RNA; there are many types of RNA in the cell, and a relative change in their number could account for the observed change in base ratio.

Alternative explanations tend to to be in terms of biochemical mechanisms for morphological theories, such as that of Hebb (1949), in which emphasis is placed on the synaptic interrelationships among neurons. These explanations tend to relate synthetic mechanisms and cell environment with the function of supplying molecules necessary for the growth (neurobiotaxis) and maintenance of peripheral connections. In this regard, two characteristics of neurons are of interest.

First, the axonal terminal is essentially devoid of ribosomes, which are necessary for protein synthesis. Thus, the proteins involved in the neuron's growth must be synthesized near the cell body and move axoplasmically (Dingman and Sporn, 1964). This rate of movement is in the order of millimeters per day. It appears that if morphological changes in the neuron's periphery is involved in memory, adequate retention measures may require a post-learning delay in the order of days or weeks, especailly if neurons with long processes are involved.

Second, there is an interrelation between a neuron's state and its transmission of impulses. For instance, in the production of synaptic transmitter substances, the synthesis of the enzymes involved is dependent upon the availability of the substrates of the enzymes. Conversely, stimulation of the neuron affects the amount of transmitter substances. Support for the latter appears to be given by the finding that brain cholinesterase level increases with learning, presumably as a reaction to an increase in the transmitter substance, acetylcholine (Rosenzweig, Krech, Bennett, and Diamond, 1962; Rosenzweig, Krech, Bennett, and Zolman, 1962).

Briggs and Kitto (1962) suggest that a cytological theory may account for the altered base ratio data. These authors suggest that learning implies neural transmission, which implies transmitter substances, which imply enzymes involved in the synthesis of these substances, which imply particular types of RNA in the synthesis of these enzymes, which imply altered base ratios reflecting an increase in these types of RNA. The administration of any substance which will affect this process may be expected to affect learning-memory, and this accounts for the data from studies involving experimental administration of RNA or RNA-related substances.

In summary of these two types of theories of memory - molecular and cytological - the former conceives of memory as represented in structural changes in a molecule and, in particular, RNA, whereas the latter conceives of memory as represented in morphological and physiological changes in the neuron, or a system of neurons.

The results of the current study are not critical to either conceptualization. Nevertheless, these results seem to be more compatible with a cytological theory, and in particular a theory which includes a temporal function in the retentive process. Such a theory might have the following characteristics.

Consider a normal neuron just before stimulation, with the transmitter substances existing at the axon in vesicles at or near the presynaptic membrane. After stimulation initiated by a learning experience, the release of the transmitter substances results in their deficit at the end of the axon, creating a gradient along the length of the axon with respect to the concentration of these substances or their constituents. As additional transmitter substances move down this gradient, a change in their concentration occurs at the cell body, and this change initiates the synthesis of those types of RNA which are necessary as templates for the synthesis of more transmitter substances. The time required for these substances to reach the axon will be dependent, in part, on axon length and on rate of impulse transmission. A higher rate implies a more rapid depletion of transmitter substances and, for a given axon, a steeper concentration gradient and, therefore, a shorter transport time; however, under the high rate, the initial reserves of transmitter substances may be depleted before replacements arrive at the axon. At normal operating rates, synaptic vesicles hold only enough transmitter substances to maintain synaptic activity for only a few minutes (Eccles, 1965).

How might arousal affect this process? It is assumed here that arousal reduces the synaptic threshold, with two factors being involved. First, the dendrites of a particular neuron may articulate with axons of a number of neurons, both excitatory and inhibitory, and the combined concentrations of their transmitter substances will determine whether the subject neuron will fire. Second, arousal is assumed to be related to activation of widespread areas of the cortex by way of the ascending fibers of the reticular formation. If all but one of the axons terminating on the subject neuron represent the arousing system, and the last axon represents the consolidating system, the threshold of the synapse to the transmission of an impulse representing the information (experience) will be reduced under high arousal conditions - this because the axons representing the arousing system will contribute most of the transmitter substances necessary, leaving the axon reflecting the consolidating system to supply less transmitter substances, which implies that it need be stimulated at a lower rate.

Lower arousal results in the recruitment of fewer neurons in the consolidating structure while the higher firing rate within this structure (not the arousing structure) results in a shorter consolidation period. Conversely, high arousal is associated with a more diffuse consolidating structure and a longer consolidation period. In both cases, the consolidating structure is tenuous during the initial phase of the consolidation period, and gains substance as consolidation proceeds. However, this initial tenuous period will be shorter under lower arousal because of the higher firing rate.

Is the above formulation consistent with the model presented in the previous section? Yes, or so it would appear. Under lower arousal, information accessibility will (1) reach the behavioral level more rapidly because of the shorter initial tenuous phase of consolidation, and (2) drop below the behavioral threshold more rapidly because of a shorter consolidation period and a smaller consolidating structure, both of which imply a less substantial engram structure. The converse is true for higher arousal. All of these characteristics are shown in Figure 4. 1.

The formulation suggested by Kleinsmith and Kaplan (1963) is rather different from the one suggested here to account for the same phenomenon. They suggest that high arousal leads to poor immediate recall because of a high firing rate which results in the relative unavailability of actively consolidating memory traces. The suggestion here is that high arousal leads to poor immediate recall because of a low firing rate which results in a longer period of time for the memory trace to gain substance.

4. 3 Comparison with Section 1 Studies

This section will involve a comparison of the current study with, first, the six basic studies cited in Section 1. 1 and then, the studies cited in Section 1. 2. As before, the six studies in Section 1. 1 will be referred to simply as Section 1. 1 studies, and the two types of studies in Section 1. 2 as Section 1. 2 studies. The comparisons will be more meaningful, however, after mention is made of four considerations: (1) the use of pre-communication measures of information, (2) the use of the term "immediate," (3) the correspondence between arousal increment and high arousal, and (4) the relation between retention and effectiveness.

Moltz and Thistlethwaite (1955) and Goldstein (1959) obtained a measure of information one week before the communication. Moltz and Thistlethwaite utilized this pre-communication information measure to separate their subjects into three levels. At each of these levels the authors tested the difference between the two treatment groups with respect to post-communication information. The manner in which Goldstein used his pre-communication is much less clear. In one place Goldstein (1957) indicates that he compared the two treatment groups with respect to both (1) immediate post-communication information minus pre-communication information, and (2) delayed post-communication information. However, elsewhere Goldstein (1959) implies that, in effect, he compared the two treatment groups with respect to both (1) immediate post-communication information, and (2) delayed post-communication information. For instance, one page 252 Goldstein (1959) states that "Differences in the effectiveness of the two appeals could not be attributed to differential recall of the content of the lectures." Hence, it is not at all clear whether Goldstein's measure of retention pertains to information or information change.

Each of the Section 1. 1 studies which obtained a retention measure, did so immediately after the communication. This measure may be referred to, in a loose sense, as an "immediate" measure of retention. However, in a strict sense, it is impossible to obtain an "immediate" measure of retention; even a retention measure obtained a few seconds after information presentation is not, strictly speaking, immediate. Therefore, it seems more appropriate to refer to such a measure as "short-term." In either case, the interval between information presentation and retention measurement should be specified. While this specification is not always clear in the Section 1. 1 studies, it appears that "short-term" may be considered as pertaining to the class period in which the communication was presented, and "long-term" pertains to one or two weeks later. The authors of the Section 1. 1 studies appear to interpret short-term retention in terms of learning, long-term retention in terms of remembering.

The current study uses the term "arousal increment" to apply to an increase in arousal over a relatively short period of time, say 1 minute, for a particular subject; the studies of Section 1. 1 use the term "high arousal" to apply to the arousal level of one group relative to a second group. There is a rough correspondence between arousal increment and high arousal: the former is more likely to occur among members of the high arousal group. (Evidence for this was presented in Section 1.31.) On this basis the current hypothesis would predict that, in a multiple-appeal study, the higher arousal appeal would show greater long-term measures of effectiveness, including retention.

Even though conformity and attitude measures of effectiveness were not obtained in the current study, it is assumed here that information retention is an important factor in influencing conformity or attitude, an assumption also made explicitly or implicitly by the Section 1.1 studies. Hence, in a rough sense, any factor which influences retention is likely also to influence conformity and attitude. It is in this sense that the three measures are referred to generically as effectiveness measures.

With these four notes out of the way, attention can be given to a comparison of the current study with the Section 1.1 studies. These comparisons will be based on several criteria, the first of which will involve the focus of the current study: retention in relation to arousal. Four of the Section 1.1 studies (Janis and Feshbach, 1953; Moltz and Thistlethwaite, 1955; Goldstein, 1959; and Berkowitz and Cottingham, 1960) made a prediction about retention in relation to arousal. For each study separately, consideration will be given below to the authors' predictions, results, and interpretations.

Janis and Feshbach (1953, page 79) predict that "The experience of being temporarily unable to terminate the disturbing affective state elicited by a discussion of a potential threat can give rise to a powerful incentive to avoid thinking or hearing about it again; this may ultimately result in failing to recall what the communicator said..." Thus, Janis and Feshbach predict that high arousal is less favorable to retention. It is not entirely clear whether Janis and Feshbach intended their prediction to apply to short-term retention only, or to retention irrespective of the time interval between the communication and the measure of retention. However, on the basis of their procedure, it would appear that Janis and Feshbach intended that their prediction apply to short-term retention only. They obtained only a short-term retention measure; their long-term (one-week) questionnaire did not include any retention items.

The prediction of Janis and Feshbach was not supported - that is, "No significant differences were found among the three experimental groups with respect to information test scores" (Janis and Feshbach, 1953, page 82). These results are interpreted by Janis and Feshbach (1953, page 89) as follows: "Our findings definitely suggest that the use of fear-arousing material of the sort presented in the illustrated talks would rarely give rise to any interference with the audience's ability to learn the content of the communication."

Moltz and Thistlethwaite (1955, page 231) predict "...that greater anxiety reduction would be associated with significantly better learning..." Thus, Moltz and Thistlethwaite predict that greater arousal reduction is more favorable to retention. It is not entirely clear whether Moltz and Thistelthwaite intended their predictions to apply to short-term retention only, or to retention irrespective of the time interval between the communication and the measure of retention. However, on the basis of their procedure, it would appear that Moltz and Thistlethwaite intended that their prediction apply to short-term retention only. They obtained only a short-term retention measure; their long-term (one-week) questionnaire did not include any retention items.

The prediction of Moltz and Thistlethwaite was not supported - that is, "...the prediction that greater anxiety reduction would be associated with significantly greater learning of factual material was not confirmed." (Moltz and Thistlethwaite, 1955, page 234). The authors do not discuss this result.

Goldstein (1957, page 8) predicts that "Copers who received the strong fear appeal will show greater retention two weeks later for the content of the lecture than will avoiders who received the same lecture. Avoiders who receive the minimal fear appeal will show greater retention two weeks later for the content of the lecture than will copers who receive the same lecture." "These hypotheses predict an interaction between personality type and level of fear arousal." (Goldstein, 1959, page 248). Thus, Goldstein predicts that high arousal is more favorable to long-term retention for copers, but less favorable for avoiders. It is not entirely clear whether Goldstein intended his prediction to apply to long-term retention only, or to retention irrespective of the time interval between the communication and the measure of retention. However, the basis of his procedure, it would appear that Goldstein intended that his prediction apply to both short-term and long-term retention. He obtained retention measures at both points in time, and with respect to each of these measures, Goldstein (1957, 1959) tested the interaction between personality type and lecture type.

The prediction of Goldstein was not supported - that is, analysis "...fails to confirm the hypothesis that copers and avoiders would retain different amounts of material from the two lectures." (Goldstein, 1957, page 28). Elsewhere, Goldstein (1959, page 250) states that "An analysis of variance of the dental information scores failed to demonstrate any significant _F_s either for the learning condition or for the scores obtained on the two-week retention test." These results are interpreted by Goldstein (1957, page 39) as follows: "The failure to demonstrate differences in the retention of the content of the two appeals by copers and avoiders suggests that there is no relationship between the ability to remember the content of a lecture and conformity to that lecture. Retention of the lecture may be necessary condition for conformity, but not a sufficient one." These two sentences appear to be contradictory. To me, the first seems to say "retention of content is not related to communication effectiveness," whereas the second seems to say "retention of content is a necessary condition for communication effectiveness."

Berkowitz and Cottingham (1960, page 41) predict "... that the frustrating experience of listening to a boring speech [Minimal group] may result in greater inattentiveness to the lecture," and on this basis they predict that low arousal will be less favorable to retention. Since these authors administered only an immediate questionnaire, it is apparent that they intended their prediction to apply to short-term retention. There is no indication, however, of whether they intended their prediction to be restricted to short-term retention.

The prediction of Berkowitz and Cottingham was not supported - that is, "There is no evidence of differences in the learning of the lecture material." (Berkowitz and Cottingham, 1960, page 41). The authors do not discuss this result.

In summary, four of the studies cited in Section 1. 1 made four different predictions with respect to retention in relation to arousal, but not one of these predictions was supported. Failure to find support for predictions is not uncommon. What seems odd, however, is the fact that not one of these four studies reevaluated the hypothesis on which the unsupported prediction was based.

Of these four studies, two (Janis and Feshbach, 1953; Goldstein, 1959) interpret gratuitously their nonsignificant results. Finding their formulations ineffective with empirical data, the authors of these two studies regress to hypothetical data.

Janis and Feshbach suggest that even though arousal was not found to have an effect on retention in the classroom, arousal may have an effect on retention in the bathroom at the time the student brushes his teeth, with high arousal less favorable to retention. Goldstein suggests that even though arousal was not found to have an effect on retention in the classroom, arousal may have an effect on retention in the bathroom at the time the student brushes his teeth, with high arousal less favorable to retention for avoiders, but more favorable for copers. Janis and Feshbach conclude that high arousal leads to "defensive avoidance," which inhibits retention later when action can be taken. Goldstein concludes that high arousal leads to "defensive reaction," which, depending on the subject's personality, either inhibits or facilitates retention later when action can be taken.

The formulation of the current study would predict that during teeth brushing, each subject, irrespective of personality or treatment group, will show retention of that information presented at a time when the subject's arousal is increasing.

Thus, the three formulations lead to three different predictions for the same experiment. Perhaps someone may be motivated to determine which of these three predictions is best supported by empirical data.

To a limited extent, the various formulations may be compared with the data at hand by determining whether the current formulation better predicts the results of Section 1. 1 than do the formulations of these studies. Unfortunately, this procedure cannot be applied profitably to the retention results of three of the four studies discussed above, because each of these three studies failed to demonstrate a significant differential arousal (experimental) effect. Hence, in these three studies (Janis and Feshbach, 1953; Goldstein, 1959; Berkowitz and Cottingham, 1960) it is not clear whether to charge the nonsignificant retention results to ineffective formulations or ineffective treatments, or both.

The fourth study (Moltz and Thistelthwaite, 1955) may be used somewhat more profitably as the basis for a comparison of the current formulation with that used by Moltz and Thistlethwaite. It is recalled that these authors obtained a significant group difference in arousal reduction, and that they hypothesized that greater arousal reduction is more favorable to retention. While their failure to obtain a significant group difference in retention clearly fails to support their hypothesis, this result does not clearly support the current formulation. This latter is based on the following consideration.

The results of the current study indicate that retention is related to arousal induction, but not to arousal reduction or to arousal level. On this basis it would be unlikely that a significant group difference in arousal reduction would lead to a significant group difference in retention. Further, since the Moltz and Thistlethwaite retention measure was a short-term measure involving 12 items, it is likely that some of the item-student combinations would lead to forgetting, retention, and reminiscence events, as defined in the current study; any retention measure which combines such events is unlikely to show a differential retention which is significant. Even though these considerations make reasonable the result obtained by Moltz and Thistlethwaite, their result cannot be taken as support for the current hypothesis, for any prediction of a nonsignificant result is trivial in the sense that its achievement fails to reject an infinite number of alternative hypotheses.

As mentioned previously, the failure of three of the above four studies to show a treatment effect precludes a critical evaluation of their formulations. This problem may be circumvented by evaluating the formulations of only those studies in Section 1. 1 which showed a significant differential arousal (a significant treatment effect). As indicated in Section 1. 18 under B1, there are three such studies (Moltz and Thistlethwaite, 1955; Janis and Terwilliger, 1962; Snider, 1962). The effectiveness measure in the first study was conformity, while the last two studies involved attitude change. Because these studies didn't utilize information retention, their results can't be compared meaningfully with those of the current study. The differential effect of arousal on attitude change and information retention has already been discussed in Section 1. 31. While these considerations preclude a cross-study evaluation, they still allow an evaluation of the hypotheses of these studies.

Moltz and Thistlethwaite (1955) showed a significant arousal-reduction effect. Their arousal-reduction hypothesis predicted that greater effectiveness would be associated with larger arousal reduction. Effectiveness was not related significantly to arousal reduction. Thus, this result fails to support the arousal-reduction hypothesis.

Janis and Terwilliger (1962) showed a significant arousal-level effect. Their arousal-minimum hypothesis predicted that greater effectiveness would be associated with smaller arousal. Effectiveness was not related significantly to arousal level. Thus, this result fails to support the arousal-minimum hypothesis.

Snider (1962) showed a significant arousal-induction effect, as well as a significant arousal-reduction effect. His arousal-reduction hypothesis predicted that greater effectiveness would be associated with larger arousal reduction. Effectiveness was related significantly to arousal induction, but not to arousal reduction. Thus, this result fails to support the arousal-reduction hypothesis.

In summary of these three studies showing a significant treatment effect, not one yielded its predicted effectiveness result. Clearly, the formulations of these studies were not supported.

So far, an attempt has been made to evaluate the formulations of Section 1. 1, using as a basis the results of the Section 1. 1 studies which (1) utilized retention as an effectiveness measure, and (2) demonstrated a significant treatment effect. Neither criterion allows as critical an evaluation of formulations as is afforded by an evaluation based on significant differential effectiveness results. As indicated in Section 1. 18 under B2, such results were shown by two studies (Berkowitz and Cottingham, 1960; Snider, 1962) of Section 1. 1. Again, unfortunately, neither result pertains to information retention, and hence can not be compared with the results of the current study.

Berkowitz and Cottingham (1960) showed that the high arousal appeal was significantly more effective. Their arguments would lead to the

Janis-Feshbach arousal-minimum hypothesis that, despite the expectation that the high arousal appeal would be more interesting, the high arousal appeal would be less effective. Thus, the Berkowitz and Cottingham result fails to support the arousal-minimum hypothesis.

Snider (1962) showed a significant arousal-induction effect, with the high arousal appeals being more effective. According to Snider (1962) this result was not predicted by his arousal-reduction hypothesis. Thus, the Snider result fails to support the arousal-reduction hypothesis.

It is time to summarize. An evaluation has been made of the hypotheses used by the six studies in Section 1. 1 on the basis of their ability to predict the differential results (high arousal vs. low arousal appeals) obtained in these six studies. Three types of results were considered: information retention, effectiveness expected on the basis of significant differential arousal, and significant differential effectiveness. All six studies were involved at least once in these results, with 4 studies involved in the first type of result, 3 in the second type, and 2 in the third. Of these 9 results, not one supported the arousal-effectiveness hypotheses of these six studies.

As further indication of the lack of predictive ability of the three hypotheses (arousal-minimum, arousal-consonant, and arousal-reduction) employed by the studies of Section 1. 1, these six studies involved a total of 16 tests of significance involving differential effectiveness, and employing three effectiveness measures (information, attitude, and conformity). Of these 16 tests, not one supported any of the three hypotheses. Two of these tests were significant but in the direction opposite to that predicted.

In addition to these last two tests, the differential (High Threat vs. Low Threat) test of the "paraphrasing of arguments" measure used by Janis and Terwilliger (1962) approached significance. Presumably, such paraphrasing implies information retention, and hence, one might legitimately compare the current prediction with the Janis and Terwilliger prediction relative to this result. It is recalled that the High Threat group gave, at the end of each paragraph, statements paraphrasing the communication's arguments to a lower mean number of paragraphs. However, even if the result were significant, it would not have allowed a critical comparison of the arousal-minimum hypothesis of the Janis and Terwilliger study with the arousal-induction hypothesis of the current study, for both studies would have predicted this result - the arousal-minimum hypothesis because it assumes that High Threat is less effective, and the arousal-induction hypothesis because it assumes that High Threat results in poor immediate recall. The two hypotheses would have predicted different results for a delayed "paraphrasing of arguments" measure, with the arousal-minimum hypothesis continuing to predict a lower score for the High Threat group, and the arousal-induction hypothesis predicting a higher score for the High Threat group. Unfortunately, Janis and Terwilliger did not obtain the long-term measure.

So far the current formulation has been compared with those used in the studies of Section 1. 1 using as a basis the results of those six studies. The comparison can also be made on the basis of the results of the current study. These results were predicted by the arousal-induction hypothesis, but they would not have been predicted on the basis of any of the three hypotheses employed by the six studies of Section 1. 1. To be supported, these three hypotheses would have to predict a change in communication effectiveness over time. However, of the 16 effectiveness measures utilized in the test of these three hypotheses, only one was obtained at two points in time following the communication. This exception was the information retention measure used by Goldstein (1959).

In this singular case, the second measure was obtained in order to test the prediction that the long-term retention of the high arousal group would be inferior to that of the low arousal group. Unfortunately, Goldstein (1957) failed to obtain a differential arousal effect. However, deleting this fact from his published article, he presumes to test the hypothesis

"...that the failure of the strong appeal to receive acceptance was due to the inability of Ss to recall the recommendations at a later date..." (Goldstein, 1959, page 250). The long-term retention difference was not significant, a result which is most directly interpreted as the result of the nonsignificant differential arousal result, though a result which is interpreted by Goldstein (1959, page 251) to suggest that "...there is no relationship between the recall of the content of a propaganda appeal and its acceptance..."

The hypothesis tested by Goldstein is of interest here because it leads to a prediction opposite to that based on the arousal-induction hypothesis, which predicts that the long-term retention of the high arousal group would be superior to that of the low arousal group. Again, then, we have the opportunity to compare alternative hypotheses. Because of the reasons given in the preceding paragraph, the comparison involving the Goldstein data was not conclusive, but the comparison involving the current data supports the current hypothesis and fails to support the hypothesis tested in the Goldstein study.

It is appropriate to summarize the results of this section so far. The current hypothesis has been compared with those of the studies of Section 1. 1 in four ways: (1) support of the Section 1. 1 hypotheses by the Section 1. 1 results, (2) support of the current hypothesis by the Section 1. 1 results, (3) support of the current hypothesis by the current results, and (4) support of the Section 1. 1 hypotheses by the current results. All hypotheses pertain to differential effectiveness in relation to differential arousal. It is assumed that the relationship between effectiveness and arousal exists between subjects and within a subject, allowing the studies of Section 1. 1 (between-subjects) to be compared meaningfully with the current study (within-subject).

In this paragraph the numbers in parenthesis correspond to those in the preceding paragraph. (1) The three hypotheses of Section 1. 1 were involved in 16 predictions, and the results in every case failed to support the associated hypothesis. (2) Two of these 16 results were significant, but since they did not pertain to information retention, they neither supported nor failed to support the current hypothesis. (3) The current arousal-induction hypothesis was involved in one prediction, and the result supported the hypothesis. (4) This result failed to support any of the hypotheses in Section 1. 1. On the basis of these results it is concluded that the predictive ability of the arousal-induction hypothesis is better than that of the arousal-reduction, arousal-minimum, and the arousal-consonant hypotheses.

In comparing the current study with those of Section 1. 2, it is more convenient to consider first the perseverative consolidation studies, and then the orienting response studies.

The results of the current study support the hypothesis of the UM studies. The current study, however, involved two additional considerations.

First, the prediction made by the UM studies pertaining to the relation of small and large arousal increments with, respectively, forgetting and reminiscence events was extended in the current study to include the prediction that medium arousal increments would be related to retention events. The predicted results emerged. A model which incorporates these results is offered in Section 4. 1.

Second, consideration was given to the temporal relation of arousal increment to information presentation. While the UM studies apparently considered only arousal increments following the onset of the stimulus, the current study considered arousal increments occurring before and after. The results suggested that arousal-induction before information presentation is more likely to be associated with a retention event, whereas arousal-induction after information presentation is more likely to be associated with a reminiscence event.

In addition to these two major considerations, other differences between the current study and the UM studies add to the delineation of the UM phenomenon (interaction between event type and resistance decrement). One difference involves the interval over which the resistance increment is considered. The UM studies used +4 seconds, whereas the current study used ±10 to ±60 seconds. The phenomenon emerged in all cases, but in the current study, the phenomenon was most prominent at about ±30 seconds. Although not described in this report, the current study also used ±4 seconds, and the phenomenon again appeared, but not as clearly as at ±30 seconds. The difference between the UM interval and the longer interval found to be optimal in the current study may reflect in part the difference in learning materials: the UM studies involved unchanging learning items separated by filler items, whereas the current study involved a continuously-changing presentation. If the rate of change of resistance is proportional to the rate of change of stimulation, one would expect a resistance decrease to be more rapid in the UM studies than in the current study. If so, this could explain the difference between the decrement intervals in the UM and current study. Further, in the current study, there was no attempt to begin the resistance decrement interval at a stimulus change point; this is one of the reasons for the reluctance in the current study to identify resistance decrement as a GSR.

A second difference involves the method of measuring retention. The UM studies used the recall method, while the current study used the recognition method. It was not expected that this difference would have a substantial influence on the emergence of the phenomenon. This expectation seems to have been supported by the fact that the phenomenon emerged in both the current and UM studies. This result argues against the suggestion made by Goldstein (1957, page 39) that his failure to show differential (high vs. low arousal) retention may be a reflection of the fact that he measured retention by the recognition method, instead of the recall method.

Now let us compare the current study with the UCLA studies. These latter studies are focussed on learning rather than memory, and they generally involve criterion measures other than retention. Thus, the current study does not articulate as readily with the UCLA studies as it does with the UM studies. The implications which the UCLA and current studies have for each other are more readily discussed after a presentation of two major conclusions to emerge from these studies and which are of interest here.

First, High and Low Orienters tend to show a consistency in response to verbal stimuli. In one study (Raskin, 1963), stimuli were presented to the subject by earphones. After the presentation of an adaptation list of 15 different filler words designed to extinguish the OR to words, a conditioned stimulus (CS) word was presented a number of times. Each CS word was followed in 10 seconds by one second of noise, and each conditioning trial (CS word + noise) was separated by filler words. The GSR to the first noise burst was taken as the measure of OR, allowing a division of the subjects into High Orienters and Low Orienters. High Orienters, relative to Low Orienters, yielded a larger mean GSR on each conditioning trial. Thus, a subject who showed a large (small) GSR to the first US tended to show a large (small) GSR to each CS, even when the latter preceded the former, as was the case on the first trial. This response consistency occurs even though the orienting levels are experimentally induced.

Second, OR is related positively to verbal learning. In the Raskin (1963) study, the above procedure was supplemented with a measure of "awareness," as indicated by the extent to which the subject reported, after the experiment was over, the experimental contingencies. Both High and Low Orienters showed awareness of the fact that the CS word was followed by the US. However, when the procedure described in the previous paragraph was followed by a number of generalization trials consisting of words semantically similar to the CS word, High Orienters, relative to Low Orienters, apparently showed (1) greater awareness of these less obvious experimental contingencies, and (2) better semantic generalization, as measured by the GSR to the generalization words.

Two other UCLA studies, these utilizing paired-associates, also found a positive relation between OR and verbal learning. Both of these studies (Belloni, 1964; Nies, 1964) were similar to a study by Standish and Champion (1960). The UCLA studies measured OR by the GSR to a word (Belloni, 1964) or a noise burst (Nies, 1964). Then, from several minutes to several weeks later, the subject learned first an easy paired-associates list, and then a difficult paired-associates list. Each list contained 10 pairs of words taken from the Kent-Rosanoff list, with the 10 stimulus words in the easy list identical to the 10 stimulus words in the difficult list. Difficulty level was manipulated by choosing for the easy list, response words most frequently given in free association, and for the difficult list, response words not given in free association. The stimulus word was presented for essentially 4 seconds, during which time the response word was anticipated, after which both words were presented for 4 seconds. The two lists, and the procedures to this point, were the same as those utilized by Standish and Champion (1960). A trial of 10 pairs was separated from the next trial by 30 seconds for the Nies study, with only an 8-second interval after every 3 trials being used in the Belloni study. Trials were continued until 2 successive errorless trials were achieved, after which there were 5 (Nies, 1964) or 8 (Belloni, 1964) further trials. Learning at each trial was measured by the latency of the correct response (arbitrarily limited to 4 seconds) averaged over subjects and over the 10 responses from each subject.

The Belloni study revealed for each list a nonsignificant learning difference in favor of the High OR group. However, for each list the OR-by-Sex interaction indicated better learning for High OR males and Low OR females, with this interaction being significant for the difficult list. High OR males showed significantly better learning than Low OR males on the difficult list, but not on the easy list. Apparently, High OR females did not differ significantly from Low OR females on either list. The Nies study, which involved males only, revealed for each list a significant learning difference in favor of the High OR group.

Thus, it appears that the OR to words (as well as to other stimuli not described here) is fairly consistent, and that the OR is related positively to verbal learning, at least for difficult tasks and males. Maltzman and Raskin (1965) emphasize the relation of OR to learning, whereas Sokolov (1963) places greater emphasis on the relation of OR to perception. Of course, response latency might reflect either perception or learning, or both, and in this regard is much like a retention score, which might reflect learning or remembering, or both. This point deserves further comment.

Even though the UCLA studies relate GSR to learning it would appear that the UCLA measures of "learning," including the physiological measures obtained in the classical conditioning and generalization study, were not entirely uncontaminated by memory effects.

The possible confounding effects of memory in learning data is suggested by a comparison of the results of Standish and Champion (1960) with those of Belloni (1964). These two studies were similar precedurally (Belloni, 1964), with the exception that the former involved a 30-second rest after each trial, whereas the latter involved an 8-second rest after three trials. The results of the current study would lead to the prediction that the Standish and Champion procedure would result in greater intertrial consolidation, which should be manifested as a greater trial-to-trial increase in learning. For the measure of learning and the 10 trials described above the Standish and Champion study showed, as predicted, larger increases in learning on successive trials for both the easy and difficult lists. The Nies (1964) study could not be readily compared because it involved only 7 trials, utilized males only, and required the subject to engage in a grip task while learning.

Having discussed the current study, first, in relation to the UM studies, and, second, in relation to the UCLA studies, it is now appropriate to consider together the results of all three; perhaps the relation of resistance decrement to learning and memory will emerge more clearly.

It may be that a particular GSR near the time of information presentation is related only to learning (UCLA) or remembering (UM), to information processing (UCLA) or information storing (UM), to attention (UCLA) or consolidation (UM). More likely, however, is the possibility that a given GSR reflects both a learning process and a remembering process. Possibly a sudden decrease in resistance reflects more a learning process, while a slower decrease reflects more a remembering process. Under this assumption, and assuming further that the rate of change of resistance is proportional to the rate of change of stimulation, then rapidly-changing stimuli may affect learning more than memory. On this basis, it may be that UCLA, UM, and current studies are ordered such that resistance decrement reflects most strongly a learning process in the former, least strongly in the latter, whereas resistance decrement reflects least strongly a remembering process in the former, most strongly in the latter. Such a formulation might explain why high orienters, relative to low orienters, show better learning, whereas high arousal subjects, relative to low arousal subjects, show poorer immediate retention.

While the above formulation is speculative, it is consistent with the formulation advanced in this study to account for the results of Table 3. 4. It was suggested there that a resistance decrement which follows information presentation is more likely to reflect a remembering process. As the decrement moves forward in time, relative to presentation, and in particular, as the decrement approaches coincidence with presentation, the decrement is more likely to reflect a learning process. Conceivably, a decrement which precedes information presentation is more likely to reflect a perceptual process.

Whether one assumes that a GSR near the time of information presentation is related to learning or to remembering, the UCLA results may appear to be inconsistent with the short-term UM results. This problem is presented more readily in the context of a resistance decrement immediately following the presentation of the stimulus word. Under the assumption that this decrement represents a response to this stimulus, it is appropriate to refer to the decrement as a GSR; this terminological convention is consistently used throughout this report and is consistent with the UCLA and UM usage of the term "GSR." The UCLA studies suggest that a larger GSR to the stimulus word is associated with higher learning of the paired associate; the UM studies demonstrate that the same GSR is associated with lower short-term recall of the paired associate and higher long-term recall. Inasmuch as the UCLA results appear to be incompatible with the short-term UM results, the matter deserves some attention.

This problem will be discussed after noting first that there is one case in which the UCLA and UM studies involve similar methods, and that in this case the studies yield similar results. One UCLA study (Belloni, 1964) and one Michigan study (Kleinsmith, Kaplan, and Tarte, 1963) obtained one measure of conductance from each subject. For simplicity, these two studies will be referred to here as the UCLA study and the UM study. Since each of these two studies also used paired-associate words, involved similar subjects, and obtained criterion measures over time, a comparison of their results would seem to be meaningful.

The UCLA study involved 120 subjects, 60 males and 60 females, while the UM study involved 64 subjects, 32 males and 32 females, with all subjects being undergraduates enrolled in introductory psychology courses.

The UCLA study obtained its conductance measure before the first learning trial, while the UM study obtained its conductance measure 30 seconds after the beginning of the learning trial. Presumably, the results are essentially independent of the point in time at which the conductance measures were obtained, for conductance typically shows a high stability reliability (e.g., Berry, 1962).

The UCLA study involved two paired-associates lists described previously, while the UM study involved a single list consisting of the final 30 paired items from a list by Melton and Safier, published by Hilgard (1951), with degrees of similarity between stimulus and response words ranging from 1.26 to .01. The difficulty level of most of the UM pairs probably falls between the two UCLA lists.

The UCLA study presented the lists in the manner described previously, while the UM study differed somewhat from that described previously for another UM study. In the UM study now being discussed, the student was told that he would have 10 seconds to study each pair, 10 seconds between pairs, six minutes between the learning and recall sessions, and 10 seconds to respond to each stimulus in the recall sessions, with only one trial in each session. These directions were followed with one group. However, subjects in a second group were told after the learning session that despite the previous instructions, they would not be tested - this in order to preclude rehearsal - and were told to return for another experiment in a week, at which time they were tested for recall. The paired-associate list was presented in one random order for all subjects during learning, and in a second random order during recall.

The UCLA study utilized mean response latency, described previously, as its criterion measure, while the UM study involved the number of items recalled correctly. The UCLA study interprets its criterion measure as one of learning, while the UM study interprets its criterion measure as one of recall, but this latter distinction need not concern us here. In the UCLA study, the last 10 trials are the only trials for which criterion scores are presented for each list.

The UCLA study showed a significant conductance-by-trial interaction for each list, with high conductance subjects showing lower learning on the first criterion trial and higher learning on the tenth trial. The cross-over appeared to occur, for the average subject, at about 10 minutes after the first presentation of the list. A similar result emerged in the UM study; at six minutes after the first (and only) presentation of the list, the high and low conductance subjects showed about the same recall, but a week later, the high conductance subjects showed a significantly higher recall.

Thus, in the one case in which the UCLA and UM studies were sufficiently similar in method as to allow a comparison of results, the results are quite similar. Both results indicate that on the basis of conductance, measured once, high conductance subjects yielded lower short-term criterion scores and higher long-term criterion scores. It is assumed here that in each study, a criterion score reflects both learning and remembering processes.

While it was possible to compare these two studies using conductance as the physiological variable, it would have been more desirable to compare them using GSR as the physiological variable, for it is this latter variable which is more similar to the physiological measure used in the current study. Unfortunately, the UCLA and UM studies cannot be compared meaningfully using GSR, for the UCLA studies measured GSR only once, before the experimental session, whereas the basic UM studies (Kleinsmith and Kaplan, 1963, 1964) measured GSR many times, immediately after the presentation of each stimulus word.

One might attempt to standardized these two procedures by either (1) modifying the UCLA procedure by assigning the one GSR value to each of the stimulus words, or (2) modifying the Michigan procedure by averaging the GSRs from each of the stimulus words. The objective of the first procedure would be to compare the UCLA results against the primary UM result (GSR-time interaction with respect to retention), while the objective of the second procedure would be to compare the UM results against the primary UCLA result (GSR-learning correlation).

The first procedure would not fulfill its objective, for the UM phenomenon is assumed to emerge only if GSR differs for successive paired associates or other learning materials. The second procedure would fulfill its objective. Presumably, the UM subject who shows a larger average GSR would have a higher learning score. Unfortunately, this hypothesis cannot be tested here because the published UM data are not in a form amenable to such a test. However, it should be noted that this hypothesis is not inconsistent with the UM phenomenon - that is, a subject with a larger (relative to other subjects) average GSR could show better (relative to other subjects) learning, as well as poorer short-term retention and better long-term retention of those items (relative to other items) for which he showed larger (relative to other items) GSRs. In fact, the argument being presented here would lead to both predictions, the between-subjects learning effect and the within-subject memory effect over time. If this prediction is supported by empirical data, the argument here would explain the apparent contradiction, discussed over these past few pages, between the UCLA results and the short-term UM results.

A test of this prediction would require a study which combines the UCLA procedure and the UM procedure. GSR to a standard stimulus would be measured before the subject is presented the paired-associates list. During the single presentation of this learning list, GSR to each stimulus word would be measured. Learning and short-term retention could be measured by, say, the percentage of correct responses on the testing list (same as learning list but response words not presented) presented immediately after the learning list. The testing list would be presented again at a later time, resulting in a long-term measure of retention.

In this procedure, percentage of correct responses is taken as a measure of both learning and memory. As has been assumed throughout this report the identification of a retention score with either learning or memory is rather arbitrary. In either case, the score would be expected to be a function of the experimental conditions. Under similar conditions, one would expect similar scores, however labelled. This expectation is fulfilled in a comparison of Belloni (1964) with Kleinsmith and Kaplan (1963), both studies involving a measure of percentage of correct responses to paired associates. The UCLA study interpreted this measure as one of learning, the UM study as one of recall. Before presenting the results of the two studies, the similarity in their methods should be noted.

Each study involved a paired-associates list, and its presentation was similar in both studies. The difficult UCLA list contained 10 0% association value pairs, while the UM list contained 8 0% association value pairs; the easy UCLA list, with its high association values, is not considered here because of its extremely dissimularity with the UM list. In both studies the stimulus word was presented for 4 seconds, followed by 4 seconds during which both words were presented. In both studies oral responses were obtained by the serial anticipation method. In the UCLA study paired associates were presented successively, while in the UM study paired associates were separated by a simple task requiring oral responses and lasting 8 seconds. In the UCLA study the list was presented repeatedly until 2 errorless and 8 subsequent trials were achieved; in the Michigan study the list was presented once during training and once during testing.

The criterion measure in each study considered here was the mean percentage of paired associates responded to correctly on each trial, with the mean taken over subjects.

The subjects in both studies were undergraduate students of both sexes enrolled in an introductory psychology course. The UCLA study involved 120 subjects, while the Michigan study involved 48 subjects.

The UCLA subject is faced with essentially the same task on the nth trial as the UM subject is on the testing (recall) trial. That is, each subject is presented a stimulus word to which he anticipates orally the response word. The fact that the trial is called a learning trial in the UCLA study and a testing trial in the UM is irrelevant. The primary distinction between the two studies is that the UCLA subject has had n-1 exposures to the list, whereas the UM subject has had only one. For n=2, since both students have had the same exposure to the list, and since relative to the beginning of the experiment, the second trial in each study occurred at about the same point in time, one would expect both students to perform about equally well. This, in fact, is what occurred.

On the second trial, the UCLA students gave correct responses to about 31% of the paired associates (Belloni, 1964, page 35), while the UM students gave about 27% (Kleinsmith and Kaplan, 1963, page 191), a difference which is not significant. A second UM study (Kleinsmith and Kaplan, 1964), utilizing a similar method, showed that the students gave correct responses to about 28% of the paired associates. Thus, at least in these cases, the scores reflect a similarity in procedure rather than a difference in label.

4.4 General Implications of the Results

The implications discussed in this section are based on the results of this study. The primary results are summarized in the next paragraph.

Consider a short interval of time near a point in time at which information is presented. The results of this study indicate that (1) a small decrement in skin resistance during the interval tends to be associated only with short-term retention, (2) a medium decrement, especially one which precedes information presentation, tends to be associated with both short- and long-term retention, and (3) a large decrement, especially one which succeeds information presentation, tends to be associated only with long-term retention.

In the implications listed below, the term "arousal increment" will be used instead of "resistance decrement."

Arousal-Induction Hypothesis

The results suggest an arousal-induction hypothesis. In terms of the communication, the arousal-induction hypothesis leads to the prediction that, other things being equal, communication effectiveness will be related positively to the number and magnitude of the arousal increments which are mediated by the communication. In terms of the communicatee, the arousal-induction hypothesis leads to the prediction that, other things being equal, long-term retention will be related positively to the number and magnitude of the arousal increments which occur during exposure to the communication.

The arousal-induction hypothesis is in fundamental opposition to hypotheses which hold that learning is not related to arousal (Becker, 1964; Janis and Feshbach, 1953) or that learning is reduced by arousal induction (Hovland, Janis, and Kelley, 1953). The arousal-induction hypothesis is either contrary to, or different from , each of the hypotheses of the studies in Section 1. 1. The arousal-induction hypothesis leads to predictions which differ to such a great extent from predictions based on any one of the Section 1. 1 hypotheses that it would be difficult to design a study in the area of arousal-communication for which the arousal-induction hypothesis and any one of the Section 1. 1 hypotheses would not predict differential results.

The results of the current study, taken with those of the UCLA and UM studies, suggest that arousal induction affects both learning and remembering processes.

Generality of Arousal Increment

This study implies that the effect of arousal increment on learning and remembering is independent of the mediator of arousal increment. No attempt was made in this study to identify these mediators, which may have been thoughts, film material, change in student position, etc. The study made no distinction between an arousal increment whose mediator is known (or presumed to be known) and an arousal increment whose mediator is not known. The fact that the UM phenomenon emerged in this study even when arousal increment preceded information presentation is further evidence in favor of this point.

In particular, the results imply that the arousal increment which enhances the learning and remembering of information A need not be mediated by information A.

If the mediator of arousal increment is irrelevant, it follows that the effect of arousal increment on learning and remembering is independent of the emotion associated with the arousal increment.

This implies that fear and joy, for instance, would both enhance learning and remembering. This implication is derived from the fact that even though different sections of the film (carnival sequence, collision sequence, etc.,) might be expected to mediate different emotions, the results were found to hold for items pertaining to each of these sequences.

The assumption of the generality of arousal increment in affecting learning and remembering implies that the preoccupation of the studies in Section 1. 1 with anxiety and fear is unnecessarily restricting. In particular, this study implies that the distinction which Janis and Feshbach (1953) make between the fear and other emotions in relation to communication effectiveness should be demonstrated, not simply alleged on the basis of psychoanalytic argument.

The generality hypothesis suggested here is readily amenable to disproof. Hopefully, attempts to disprove it will be based on empirical data.

Flexibility of Arousal Increment to Information Presentation

The study implies that long-term retention of information is enhanced if either (1) a large arousal increment occurs near the point in time of information presentation, or (2) information is presented near the point in time at which a large arousal increment occurs.

The only reasons for listing both possibilities is to emphasize the two basic alternatives available to the communicator (teacher, textbook writer, educational film producer, etc.). He may present critical information at about the same time as "arousal-inducing" material. Alternatively, he may monitor student arousal, presenting critical information only during arousal induction. The former method is by far the more commonly used, being required in those situations in which mass communications and group instruction are employed. The latter method is more applicable to individualized adaptive instruction, in which the information presentation is determined by a heuristic program one of whose inputs is the results of the real-time analysis of the student's arousal, as indicated by, say, skin resistance.

Limitation of Short-Term Measurement

Taken together, the results of the current and UCLA studies imply that a short-term retention measure of learning fails to indicate as great a superiority for high orienters as would be indicated by a long-term retention measure. A similar statement could be made for two treatments which affect arousal differentially. For example, if the learning effects of two such treatments, A and B, are measured by retention immediately after the experiences, and if A shows more learning because of a greater arousal induction, then it is assumed here that the superiority of A would be even more marked at a later time.

Given this interpretation, it appears that the common procedure in both practice and research of obtaining retention measures of learning at the end of the learning session leads to results which need to be qualified by the point in time at which the learning measure is obtained. Statements pertaining to learning in relation to arousal which are not so qualified, such as those of Janis and Feshbach (1953), should be eschewed.

4.5 Implications for Traffic Safety Films

The implications to be discussed in this section are restricted to those traffic safety films which attempt to convey information, the long-term retention of which has been shown to affect driving behavior. Unfortunately, little is known about the kind of knowledge which leads to safer driving. For example, does knowledge of stopping distances for various speeds lead to safer driving? Does a knowledge of laws of motion lead to more appropriate driving behavior in emergency situations? Answers to such questions are essential to the preparation of effective traffic safety films. It is assumed here that these answers will be available before the results of this study are applied to the preparation of traffic safety films.

This section involves a second assumption. Many traffic safety films are of the attitudinal type having as their primary purpose the development or modification of attitudes assumed to be related to safer driving. Even if it were true that these attitudes are known, and even if it were true that their change would lead to safer driving, a problem which remains is that of changing attitudes - and in the desired direction. The extent of this problem is shown by the fact that of 15 studies involving film-mediated attitude change, over half failed to find a significant change in the direction intended by the communicator, with four of these showing a significant change in the direction opposite to that intended by the communicator (Levonian, 1963a). Such an opposite effect has been referred to as a "boomerang effect" by Hovland, Lumsdaine, and Sheffield (1949), an effect from which traffic safety communications are not immune (Creative Research Associates, 1961; Naisbitt, 1961).

Even though it is not the primary purpose of an attitudinal film to transmit information, such films generally include information under the assumption that one way to change an attitude is to change its informational base. On this assumption, the implications of the current study are directed as much to the informational aspects of attitudinal films as to informational films. However, because of the difference in information and attitude, measured short-term, a difference discussed in Sections 1.16 and 1.31, the implications of this study do not apply in any direct way to the non-informational aspects of attitudinal films. An additional consideration which applies to the attitudinal film derives from the fact that the presentation of information relevant to the attitude is in itself arousal inducing (McGinnies and Aiba, 1965; Stern, Winokur, Graham, and Graham, 1961).

Assuming, then, that we know what information will lead to safer driving, whether this information influences driving behavior by way of the driver's cognitive system or by the way of his attitudinal system, we will focus here on the implications of the results for the problem of utilizing films to convey this information to the driver or prospective driver.

First, the results imply a caution in accepting the position that traffic safety films are less effective if they induce anxiety (Malfetti, 1961; Merrill, 1962). Again, it is emphasized that this implication is advanced under the assumption that the relationship between arousal induction and learning-remembering which emerged in this study is independent of the emotion associated with the arousal. On the other hand the results do not imply that a film viewed under anxiety-inducing conditions is necessarily as effective as the same film viewed under conditions involving the same arousal pattern which, however, reflects a different emotion - pleasure, say. It is assumed here that anxiety which is mediated by a film and which generalizes to certain traffic situation may result in less appropriate driving in certain traffic situation, as well as in more appropriate driving in other traffic situations. The role of anxiety in driving should be elucidated by experimental inquiry, not fiat.

The point made here is that an anxiety-inducing film, per se, is not necessarily less effective. The results should not be interpreted as implying that anxiety induction necessarily improves communication effectiveness; the results should be interpreted as implying that anxiety induction does not necessarily reduce communication effectiveness.

Second, the results imply that retention of traffic safety film information is related only indirectly to the arousal level mediated by the film, at least when the term "arousal level" is used as it was in Section 1. 1 studies to distinguish between high and low arousal communications. This point derives from the fact that the current results revealed that information retention (1) is related to arousal induction, and (2) is not related to arousal level. Nevertheless, a high arousal film may be effective by virtue of increasing the possibility that critical information will be presented during periods of arousal induction.

This argument leads to a recommendation which is in direct opposition to that of Janis and Feshbach (1953), who recommend that an arousal-inducing communication should incorporate its recommendations at the end, after arousal induction. The current results lead to the recommendations at the beginning, during arousal induction. The issue is clearly delineated and may be tested directly. The results of one experiment by Schlesinger, Fischer, and Cohen (1965) is consistent with the hypothesis suggested here, though the hypothesis was not critically tested because the time separation between arousal induction and information presentation in that experiment was greater than the optimal interval which emerged in the current study.

Third, the results imply that retention of traffic safety film information is related only indirectly to the arousal reduction mediated by the film. The argument here has the same basis as that in the argument above. That is, an arousal-reducing film may be effective by virtue of implying preceding arousal induction, as well as of implying the possibility for subsequent arousal induction.

Fourth, the results imply that retention of traffic safety film information is related to film-mediated changes in arousal level, and to the synchronization of critical film information with the arousal-inducing phases. This suggests the alternation of arousal-inducing and arousal-reducing sequences in a traffic safety film, with critical information presented only during the arousal-inducing sequences. This is probably the most practical implication of this study.

For studies in which physiological measurement of arousal is precluded, a crude measure of arousal induction and information synchronization may be taken as the ratio of reminiscence to forgetting events: the higher the ratio the more effective the film.

Fifth, the results imply that arousal need not be mediated by content related to the critical information. Presumable, the retention of film information pertaining to, say, safety belts, is related to the synchronization of information presentation with arousal induction but independent of whether the arousal was induced by an automobile collision or by an exciting ski scene. Traffic safety films are generally prepared on the basis of cinematic characteristics, such as continuity of action, as well as on the relevance of sound to picture, and vice versa. The current study suggests that such practices may not enhance film effectiveness, and in that sense may be unduly restrictive.

Sixth, the hetereogeneity of autonomic activity found in this study implies a severe problem associated with the preparation and utilization of traffic safety films. If arousal induction is to be under film control, as has been assumed in this section, the film maker is faced with the problem of employing content which induces arousal in all communicatees. He may employ any radical change in picture or sound, such as a flash of light or a loud noise. Such unconditioned stimuli are likely to be more effective than conditioned stimuli, but if the latter are used, an attempt should be made to use those which are culture-common. We may expect this attempt to be increasingly successful with an increase in the homogeneity of the cultural group. Thus, a film made for driver education students has the potentiality of being more effective than one made for a wider audience but used in driver education. Similarly, a film made for driver education students in a particular school district has the potentiality of being more effective than a film made for driver education students in general but used in that school district.

As a corollary, traffic safety material presented in a purely visual medium, such as print, will not have the potentiality of being as effective as material which allows sound synchronization. For instance, it would be predicted on the basis of the current results that programed traffic safety material presented via a book would not be as effective as the same material presented via a machine which made a clunk on the presentation of each frame, with both conditions requiring essentially the same motor response on the part of the student. Further, it would be predicted that oiling the machine would eliminate the superiority for the machine.

As another corollary, an auditory medium, such as tape, will have a greater effectiveness potential than a non-moving visual medium such as print - this by virtue of the greater ease in eliciting an unconditioned response by auditory stimuli than by non-moving visual stimuli.

Seventh, inasmuch as the OR habituates (Seward and Seward, 1935), the time interval between the US and information presentation needs to be shortened as the film preceeds if it is intended that the latter material be retained to the same level as the earlier material. Alternatively, if the same US is used throughout the film, and if the interval is not shortened, material toward the end of the film will not be retained to the same level as earlier material.

Eighth, a test to assess long-term remembering of traffic safety film information should not be administered during the same period in which the film is shown.

Summary

Six basic studies pertaining to communication effectiveness in relation to anxiety arousal involved 16 measures of effectiveness to test three hypotheses: that effectiveness is enhanced if the communication mediates (1) a low level of arousal, (2) an arousal level consonant with that level at which the communicatee learns best, or (3) an arousal reduction. Not one of these 16 tests supported these three hypotheses. However, research in the area of learning and remembering indicates that a larger GSR at the time of information presentation is associated with greater learning and long-term retention of that information.

This review of the literature suggested the following hypothesis for the current study: forgetting, retention, and reminiscence events, respectively, are associated with smaller, medium, and larger increases in arousal (arousal increments) near the time of information presentation. An information test was administered 10 minutes after information presentation as well as a week later, and a forgetting event was defined as one for which the subject gave the correct answer the first time but not the second, a retention event involved correct answers both times, while a reminiscence event involved a correct answer the second time but not the first. Arousal increment was measured by resistance decrement.

On the day that a traffic safety film was scheduled for showing in four high school driver education classes, skin resistance measures were obtained from each of 83 subjects during class showing of the film. The recording of these measures of arousal utilized a system designed especially for classroom use. A 15-item questionnaire was administered immediately after the showing, as well as one week later. Each item pertained to either auditory or visual information presented at a specific point in the film. With this as the midpoint of a 1-minute interval, arousal increment was defined as the maximum resistance decrement within the interval.

Of the 30 subjects who showed at least one forgetting, retention, and reminiscence event, respectively, 11 subjects showed smaller, medium, and larger arousal increments near the time of presentation of information associated with these events. This finding was significant at the .01 level, and supports the primary hypothesis of the study. A second finding pertained to the presentation-arousal order: there was a tendency for arousal induction to precede presentation for retention events, and for arousal induction to follow presentation for reminiscence events.

There results suggested a model relating retention interval and arousal increment to information accessibility, where information accessibility is defined here as the extent to which information which is stored, or being stored, in the subject is available to the subject. A possible biological basis for the results was also discussed.

With respect to traffic safety films, the results imply (1) that caution should be exercised in accepting the position that traffic safety films are less effective if they induce anxiety, (2) that an effective film is likely to be one which alternates arousal-inducing and arousal-reducing sequences, with information presented only during the arousal-inducing sequences, (3) that arousal induction need not be mediated by content related to the information to be retained, (4) that the inter-subject heterogeneity of autonomic activity presents a severe prblem for the film maker in the selection of film content which will induce arousal in each subject, and (5) that a test to assess long-term remembering of information should not be administered during the period in which the film is shown.

References

1. Aas, A. Mutiliation Fantasies and Autonomic Response. Oslo, Norway: Oslo University Press, 1958.
2. Air Force Cambridge Research Center, Operational Applications Laboratory. Technical Report 54-50, 1954.
3. Alexander, F., Flagg, G. R., Foster, S., Clemens, T., and Blahd, W. Experimental studies of emotional stress: 1. Hyperthyroidism. Psychosomatic Medicine, 1961, 22, 104-114.
4. Allport, Gordon W. The trend in motivational theory. Journal of Orthopsychiatry, 1953, 23, 107-119.
5. Alter, Millicent, and Silverman, Robert. The response in programed instruction. Journal of Programed Instruction, 1962, 1(1), 55-78.
6. American Insurance Association. Safety Film News, 1965, No. 18.
7. Attneave, Fred. Applications of Information Theory to Psychology. New York: Holt, 1959.
8. Bandura, Albert, Ross, Dorothea, and Ross, Sheila A. Imitation of film-mediated aggressive models. Journal of Abnormal and Social Psychology, 1963, 66(1), 3-11.
9. Bartlett, M. S. The statistical significance of odd bits of information. Biometrika, 1952, 39, 228-237.
10. Beam, J. Serial learning and conditioning under real-life stress. Journal of Abnormal and Social Psychology, 1955, 51, 543-551.
11. Becker, Samuel L. Reaction profiles; studies of methodology. Journal of Broadcasting, Summer, 1960, 253-268.
12. Becker, Samuel L. Interest, tension, and retention. Audiovisual Communication Review, 1964, 12(3), 277-291.
13. Belloni, Marigold L. The Relationship of the Orienting Reaction and Manifest Anxiety to Paired-Associates Learning. Unpublished doctor's dissertation. University of California, Los Angeles, 1964.
14. Berkowitz, Leonard. The effects of observing violence, Scientific American, 1964, 210(2), 35-41.
15. Berkowitz, Leonard, and Cottingham, Donald R. The interest value and relevance of fear arousing communications. Journal of Abnormal and Social Psychology, 1960, 60(1), 37-43.
16. Berkowitz, Leonard, and Rawlings, Edna. Effects of film violence on inhibitions against subsequent aggression. Journal of Abnormal and Social Psychology, 1963, 66(5), 405-412.

17. Berlyne, D. E. Conflict, Arousal and Curiosity. New York: McGraw-Hill, 1960.

18. Berlyne, D. E. Emotional aspects of learning. In Farnsworth, Paul R., McNemar, Olga, and McNemar, Quinn (Eds.), Annual Review of Psychology. Palo Alto, Calif.: Annual Reviews, 1964, pp 115-142.

19. Berry, R. N. Skin conductance levels and verbal recall. Journal of Experimental Psychology, 1962, 63(3), 275-277.

20. Blumenthal, Murray (Ed.) The Denver Symposium on Mass Communications Research for Safety. Chicago: National Safety Council, 1964.

21. Brady, J. V. The effect of electroconvulsive shock on a conditioned emotional response: The significance of the interval between the emotional conditioning and the electroconvulsive shock. Journal of Comparative and Physiological Psychology, 1952, 45, 9-13.

22. Briggs, Michael H., and Kitto, Barrie G. The molecular basis of memory and learning. Psychological Review, 1962, 69(6), 537-541.

23. Brody, Leon. A study of the learning of selected driving skills through exposure to a specially produced motion picture film Traffic Safety Research Review, 1961, 5(2), 25-29.

24. Burnham, W. H. Retroactive amnesia: Illustrative cases and a tentative explanation. American Journal of Psychology, 1903, 14, 382-396.

25. Burns, B. D. The production of afterbursts in isolated unanesthetized cerebral cortex. Journal of Physiology, 1954, 125, 427-446.

26. Cameron, D. Ewen. The process of remembering. British Journal of Psychiatry, 1963, 109(Whole No. 460), 325-340.

27. Cameron, D. Ewen, Sved, S., Solyom, L., Wainrib, B., and Barik, H. Effects of ribonucleic acid on memory defect in the aged. American Journal of Psychiatry, 1963, 120(4), 320-325.

28. Cattell, R. B. The significance of the actual resistances in psychogalvanic experiments. British Journal of Psychology, 1928, 19, 34-43.

29. Cook, Leonard, Davidson, Arnold B., Dixon, J. Davis, Green, Harry, and Fellows, Edwin J. Ribonucleic acid: Effect on conditioned behavior in rats. Science, 1963, 141(Whole No. 3577), 268-269.

30. Corning, W. C., and John, E. R. Effect of ribonuclease on retention of conditioned response in regenerated planarians. Science, 1961, 134, 1363-1365.

31. Creative Research Associates. *The Effects of CBS Reports "The Great Holiday Massacre" on Attitudes Toward Safety and the National Safety Council.* Chicago, 1961.

32. Davis, R. C. Response patterns. *Transactions of the New York Academy of Sciences,* Series II, 1957, 19(8), 731-739.

33. Davison, Leslie A. *Adaptation to a Threatening Stimulus.* Unpublished doctor's dissertation. University of California, Berkeley, 1963.

34. Dingman, Wesley, and Sporn, Michael B. The incorporation of B-azaguanine into rat brain RNA and its effect on maze learning by the rat: An inquiry into the biochemical basis of memory. *Journal of Psychiatric Research,* 1961, 1, 1-11.

35. Dingman, Wesley, and Sporn, Michael B. Molecular theories of memory. *Science,* 1964, 144(Whole No. 3614), 26-29.

36. Dysinger, W. S., and Ruckmick, C. A. *The Emotional Responses of Children to the Motion Picture Situation.* N. Y.: MacMillan, 1933.

37. Eccles, Sir John. The synapse. *Scientific American,* 1965, 212(1), 56-66.

38. Edwards, A. L., and English, H. B. Reminiscence in relation to differential difficulty. *Journal of Experimental Psychology,* 1939, 25, 100-108.

39. Egyhazi, E., and Hyden, H. Experimentally induced changes in the base composition of the ribonucleic acids of isolated nerve cells and oligodendroglial cells. *Journal of Biophysical and Biochemical Cytology,* 1961, 10, 403-410.

40. Eriksen, Charles W., and Wechsler, Henry. Some effects of experimentally induced anxiety upon discrimination behavior. *Journal of Abnormal and Social Psychology,* 1955, 51(3), 458-463.

41. Evans, James L., Glaser, Robert, and Homme, Lloyd E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine and R. Glaser (Eds.), *Teaching Machines and Programmed Learning.* Washington, D. C.: National Education Assn., 1960.

42. Feldhusen, John F., and Birt, Andrew. A study of nine methods of presentation of programed learning material. *Journal of Educational Research,* 1962, 55, 461-466.

43. Ferraro, Douglas P., Teal, Gilbert E., Fabrizio, Ralph A. *Your Car and Safe Driving.* New York: Macmillan, 1964.

44. Fisher, Ronald A. *Statistical Methods for Research Workers.* Fourth Edition. London: Oliver and Boyd, 1932.

45. Fisher, Ronald A. Statistical Methods for Research Workers. Twelfth Edition. London: Oliver and Boyd, 1954.

46. Fisher, Ronald A. Statistical Methods and Scientific Inference. London: Oliver and Boyd, 1956.

47. Gaito, John. A biochemical approach to learning and memory. Psychological Review, 1961, 68(4), 288-292.

48. Gaito, John. DNA and RNA as memory molecules. Psychological Review, 1963, 70(5), 471-480

49. Gaito, John, and Zavala, Albert, Neurochemistry and learning. Psychological Bulletin, 1964, 61(1), 45-62.

50. Garner, W. R., and McGill, W. J. The relation between information and variance analyses. Psychometrika, 1956, 21, 219-228.

51. Gestant, H. J., and Bert, J. EEG changes during cinematographic presentation. Electoencephalography and Clinical Neurophysiology, 1954, 6, 433-444.

52. Glickman, Stephen E. Deficits in avoidance learning produced by stimulation of the ascending reticular formation. Canadian Journal of Psychology, 1958, 12, 97-102.

53. Glickman, Stephen E. Perseverative neural processes and consolidation of the memory trace. Psychological Bulletin, 1961, 58(3), 218-233.

54. Goldbeck, Robert A., Campbell, Vincent N., and Llewellyn, Joan E. Further experimental evidence on response modes in automated instruction. Technical Report No. 3. Pittsburgh: American Institute for Research, 1960.

55. Goldberg, Alfred L. Memory mechanism. Science, 1964, 144 (Whole No. 3626), 1529.

56. Goldstein, Michael J. The Relationship Between Coping and Avoiding Behavior and Attitude Modification. Unpublished doctor's dissertation. University of Washington, 1957.

57. Goldstein, Michael J. The relationship between coping and avoiding behavior and response to fear-arousing propaganda. Journal of Social and Abnormal Psychology, 1959, 58(2), 247-252.

58. Goldstein, Michael J. The social desirability variable in attitude research. Journal of Social Psychology, 1960, 52(1), 103-108.

59. Gordon, Mordecai H., Loveland, Edward H., and Cureton, Edward E. An extended table of chi-square for two degrees of freedom, for use in combining probabilities from independent samples. Psychometrika, 1952, 17(3), 311-316.

60. Graham, David T., and Graham, Frances K. Specific Relations of Attitudes to Physiological Change. Progress Report, July 1, 1964, PHS Grant MH-02011. School of Medicine, The University of Wisconsin.

61. Guba, Egon, Wolf, Willavene, de Groot, Sybil, Knemeyer, Manfred, Van Atta, Ralph, Light, Larry. Eye movements and TV viewing in children. Audiovisual Communication Review, 1964, 12(4), 386-401.

62. Guilford, J. P. Fundamental Statistics in Psychology and Education. Third Ed. New York: McGraw-Hill, 1956.

63. Guttman, L. The basis of scalogram analysis. In S. A. Stouffer, et al., Measurement and Prediction. Princeton: Princeton University Press, 1950. pp 60-90.

64. Harris, Jean H. Attitudes and Survival. Harrisburg, Pennsylvania: Lear Siegler, 1965.

65. Hartmann, George W. A field experiment on the comparative effectiveness of "emotional" and "rational" political leaflets in determining election results. Journal of Abnormal and Social Psychology, 1936, 31(1), 99-114.

66. Hebb, D. O. The Organization of Behavior. New York: Wiley, 1949.

67. Hess, Eckhard H., and Polt, James M. Pupil size as related to interest value of visual stimuli. Science, 1960, 132 (Whole No. 3423), 349-350.

68. Hess, Eckhard H., and Polt, James M. Pupil size in relation to mental activity during simple problem-solving. Science, 1964, 143 (Whole No. 3611), 1190-1192.

69. Hilgard, Ernest R. Methods and procedures in the study of learning Chapter 15 in S. S. Stevens (Ed.), Handbook of Experimental Psychology. New York: Wiley, 1951.

70. Hovland, Carl I., Janis, Irving L., and Kelley, Harold H. Communication and Persuasion. New Haven: Yale University Press, 1953.

71. Hovland, Carl I., Lumsdaine, Arthur A., and Sheffield, Fred D. Experiments on Mass Communication. Princeton: Princeton University Press, 1949.

72. Humphrey, John H. Emotional reactions to abstract motion on film. Journal of the University Film Producers Association, 1950, 3, 11-12.

73. Hyden, H. Satellite cells in the nervous system. Scientific American, 1961, 205(6), 62-70.

74. Jackson, Philip W., and Strattner, Nina. Meaningful learning and retention: noncognitive variables. Review of Educational Research, 1964, 34(5), 513-529.

75. Janis, Irving L., and Feshbach, Seymour. Effects of fear-arousing communications. Journal of Abnormal and Social Psychology, 1953, 48(3), 78-92.

76. Janis, Irving L., and Feshbach, Seymour. Personality differences associated with responsiveness to fear-arousing communications. Journal of Personality, 1954, 23(2), 154-166.

77. Janis, Irving L., and Milholland, Harry C. The influence of threat appeals on selective learning of the content of a persuasive communication. Journal of Psychology, 1954, 37(1), 75-80.

78. Janis, Irving L., and Terwilliger, Robert F. An experimental study of psychological resistances to fear-arousing communications. Journal of Abnormal and Social Psychology, 1962, 65(6), 403-410.

79. Johnson, Palmer O., and Jackson, Robert W. B. Modern Statistical Methods: Descriptive and Inductive. Chicago: Rand McNally, 1959.

80. Kaess, Walter, and Zeaman, David. Positive and negative knowledge of results on a Pressey-type punchboard. Journal of Experimental Psychology, 1960, 60, 12-17.

81. Kanner, Joseph H., and Sulzer, Richard L. Overt and covert rehearsal of fifty per cent versus one hundred per cent of the material in film learning. In A. A. Lumsdaine (Ed.), Student Response in Programmed Instruction. Washington, D.C.: NAS-NRC, 1961.

82. Kendall, Maurice G. The Advanced Theory of Statistics, Vol. II. London: Charles Griffin, 1946.

83. Kendall, Maurice G., and Stuart, Alan. The Advanced Theory of Statistics, Vol. II. London: Charles Griffin, 1961.

84. Kleinsmith, Lewis J., and Kaplan, Stephen. Paired-associate learning as a function of arousal and interpolated interval. Journal of Experimental Psychology, 1963, 65(2), 190-193.

85. Kleinsmith, Lewis J., and Kaplan, Stephen. The interaction of arousal and recall interval in nonsense syllable paired-associate learning. Journal of Experimental Psychology, 1964, 67(2), 124-126.

86. Kleinsmith, Lewis J., Kaplan, Stephen, and Tarte, Robert D. The relationship of arousal to short- and long-term verbal recall. Canadian Journal of Psychology, 1963, 17(4), 393-397.

87. Kleitman, N. The effect of motion pictures on body temperature. Science, 1945, 101, 507-508.

88. Klemmer, E. T. Tables for computing informational measures. In Henry Quastler (Ed.), Information Theory in Psychology. Glencoe, Illinois: Free Press, 1955.

89. Knower, Franklin H. Experimental studies of changes in attitudes: 1. A study of the effect of oral argument on changes of attitude. Journal of Social Psychology, 1935, 6(3), 315-347.

90. Kornberg, A. Biologic synthesis of deoxyribonucleic acids. Science, 1961, 131, 1503-1508.

91. Krumboltz, John D., and Weisman, Ronald G. The effect of overt versus covert responding to programed instruction on immediate and delayed retention. Journal of Educational Psychology, 1962, 53(2), 89-92.

92. Kullback, Solomon. Information Theory and Statistics. New York: Wiley, 1959.

93. Lazarus, Richard S. A laboratory approach to the dynamics of psychological stress. American Psychologist, 1964, 19(6), 400-411.

94. Lazarus, Richard S., Speisman, Joseph C., Mordkoff, Arnold M., and Davison, Leslie A. A laboratory study of psychological stress produced by a motion picture film. Psychological Monographs, 1962, 76, Whole No. 553.

95. Levonian, Edward. Measurement and Analysis of Physiological Response to Film. Los Angeles: University of California. Report No. 62-66, 1962.

96. Levonian, Edward. Opinion change as mediated by an audience-tailored film. Audiovisual Communications Review, 1963a, 11(4), 104-113.

97. Levonian, Edward. Skin resistance electrode for classroom use. Psychological Reports, 1963b, 13, 451-460.

98. Lindsley, Donald B. Psychophysiology and motivation. In Jones, Marshall R. (Ed.) Nebraska Symposium on Motivation, 1957. Lincoln, Nebraska: University of Nebraska Press, 1957.

99. Lövass, O. J. Effect of exposure to symbolic aggression on aggressive behavior. Child Development, 1961, 32, 37-44.

100. Luborsky, Lester, Blinder, Barton, and Mackworth, Norman. Eye fixation and recall of pictures as a function of GSR responsivity. Perceptual and Motor Skills, 1963, 16, 469-483.

101. Maccoby, Eleanor E., Wilson, William C., and Burton, Roger V. Differential movie-viewing behavior of male and female viewers. *Journal of Personality*, 1958, 26, 259-267.

102. Malfetti, James L. Scare technique and traffic safety. *Traffic Quarterly*, 1961, 15(2), 318-330.

103. Maltzman, Irving, and Raskin, David C. Effects of individual differences in the orienting reflex on conditioning and complex processes. *Journal of Experimental Research on Personality*, 1965, 1(1), 1-16.

104. Maltzman, Irving, Eisman, Eugene, and Morrisett, Lloyd, Jr.. Rational learning under manifest and induced anxiety. *Psychological Reports*, 1961, 8, 357-366.

105. Mandler, G., and Sarason, S. B. A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 1952, 47, 166-173.

106. McAshan, Hildreth H. *An Experimental Study of Traffic Safety Films*. Unpublished doctor's dissertation. East Lansing: Michigan State University, 1960.

107. McGill, William J. Multivariate information transmission. *Psychometrika*, 1954, 19, 97-116.

108. McGill, William J. Isomorphism in statistical analysis. In Henry Quastler (Ed.), *Information Theory in Psychology*. Glencoe, Illinois: Free Press, 1955.

109. McGinnies, Elliott, and Aiba, Hitoshi. Persuasion and emotional response: a cross-cultural study. *Psychological Reports*, 1965, 16(2), 503-510.

110. McGuire, William J. Audience participation and audio-visual instruction: overt-covert responding and rate of presentation. In A. A. Lumsdaine (Ed.), *Student Response in Programmed Instruction*. Washington, D.C.: NAS-NRC, 1961.

111. Menefee, Selden C., and Granneberg, Audrey G. Propaganda and opinions on foreign policy. *Journal of Social Psychology*, 1940, 11(2), 393-404.

112. Merrill, Irving R. *Films in Driver Education*. East Lansing: Michigan State University, Highway Traffic Safety Center. Undated, but probably published in 1959.

113. Merrill, Irving R. Attitude films and attitude change. *Audiovisual Communication Review*, 1962, 10(1), 3-13.

114. Merrill, Irving R., and McAshan, Hildreth H. Predicting learning, attitude shift, and skill improvement from a traffic safety film. *Audiovisual Communication Review*, 1960, 8(6), 263-274.

115. Miller, G. A., and Ross, P. M. Tables of n $\log_2$ n and n $\log_{10}$ n for n from 1 to 1000. Tech. Rep. No. 60, Lincoln Laboratory, 1954.

116. Moltz, Howard, and Thistlethwaite, Donald L. Attitude modification and anxiety reduction. Journal of Abnormal and Social Psychology, 1955, 50(2), 231-237.

117. Mordkoff, Arnold Melvin The Relation between Physiological and Phenomenological Indices of Arousal. Unpublished doctor's dissertation. University of California, Berkeley, 1963.

118. Mussen, P. H., and Rutherford, E. Effects of aggressive cartoons on children's aggressive play. Journal of Abnormal and Social Psychology, 1961, 62, 461-464.

119. Naisbitt, John. "The Great Holiday Massacre," A study of impact. Traffic Safety, 1961, 36, 12-15.

120. National Drivers Test. New York: Random House, 1965.

121. National Education Association. Guide to Selected Safety Education Films, 1960.

122. National Safety Council. National Directory of Safety Films, 1965-1966 Edition, 1965.

123. Nies, Richard C. Orienting Response and Drive as Distinguishable Concepts in Learning Performance. Unpublished doctor's dissertation. University of California, Los Angeles, 1964.

124. Olds, James. A preliminary mapping of electrical reinforcing effects in the rat brain. Journal of Comparative and Physiological Psychology, 1956, 49, 281-285.

125. Osgood, Charles E. Method and Theory in Experimental Psychology. New York: Oxford University Press, 1953.

126. Pauling, Linus. A molecular theory of general anesthesia. Science, 1964, 134 (Whole No. 3471), 15-21.

127. Pearlman, C. A., Jr., Sharpless, S. K., and Jarvik, M. E. Retrograde amnesia produced by anesthetic and convulsant agents. Journal of Comparative and Physiological Psychology, 1961, 54, 109-112.

128. Pearson, Karl. On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. Biometrika, 1933, 25, 379-410.

129. Pribram, Karl. A review of theory in physiological psychology. In Paul R. Farnsworth and Quinn McNemar (Eds.), Annual Review of Psychology. Stanford: Annual Reviews, 1960.

130. Pribram, Karl. Neurological notes on the art of educating. In Ernest R. Hilgard (Ed.), Theories of Learning and Instruction. Chicago: University of Chicago Press, 1964.

131. Raskin, David C. Some Factors Influencing Semantic Conditioning and Generalization of Autonomic Responses. Unpublished doctor's dissertation. University of California, Los Angeles, 1963.

132. Roe, Arnold. Automated teaching methods using linear programs. Journal of Applied Psychology, 1962, 46(3), 198-201.

133. Rohloff, Joseph W. Physics and the Driver. Harrisburg, Pennsylvania: Lear Siegler, 1964.

134. Rosenstein, Alvin J. Psychometric versus physiological anxiety and serial learning. Journal of Personality, 1960, 28(3), 279-292.

135. Rosenzweig, Mark R., Krech, David, Bennett, Edward L., and Diamond, Marian C. Effects of environmental complexity and training on brain chemistry and anatomy: A replication and extension. Journal of Comparative and Physiological Psychology, 1962, 55(4), 429-437.

136. Rosenzweig, Mark R., Krech, David, Bennett, Edward L., and Zolman, James F. Variation in environmental complexity and brain measures. Journal of Comparative and Physiological Psychology, 1962, 55(6), 1092-1095.

137. Schlesinger, Lawrence E., Fischer, Edward H., and Cohen, Stanley L. The Effects of Relevant Emotional Content on Performance and Learning in Programmed Instruction. Final Report, Grant No. 7-18-0100-197, U.S. Office of Education, 1965.

138. Schwartz, B. J. An empirical test of two Freudian hypotheses concerning castration anxiety. Journal of Personality, 1956, 24, 318-327.

139. Seward, J. P., and Seward, G. H. The relation of the galvanic skin reactions to preceding resistance. Journal of Experimental Psychology, 1935, 18, 64-79.

140. Snider, Marvin. The Relationship Between Fear Arousal and Attitude Change. Unpublished doctor's dissertation. Boston University, 1962.

141. Sokolov, E. N. Perception and the Conditioned Response. New York; Macmillan, 1963.

142. Standish, R. R., and Champion, R. A. Task difficulty and drive in verbal learning. Journal of Experimental Psychology, 1960, 59(6), 361-365.

143. Stern, J. A., Winokur, G., Graham, D. T., and Graham, F. K. Alterations in physiological measures during experimentally induced attitudes. Journal of Psychosomatic Research, 1961, 5, 73-82.

144. Stolurow, L. M., and Walker, C. C. A comparison of overt and covert response in programed learning. Journal of Educational Research, 1962, 55, 421-432.

145. Velte, Joseph J. Freeway Driving. Harrisburg, Pennsylvania: Lear Siegler, 1965.

146. Villars, Donald Statlar. Statistical Design and Analysis of Experiments for Development Research. Dubuque, Iowa: Wm. C. Brown Co., 1951.

147. Walker, Edward, Memory Storage as a function of arousal and time. Paper presented at the Annual Meeting of the American Assn. for the Advancement of Science, Philadelphia, 1962.

148. Wallis, W. Allen. Compounding probabilities from independent significance tests. Econometrica, 1942, 10(3and4), 229-248.

149. Walters, Richard H., Thomas, Edward Llewellyn, Acker, C. William. Enhancement of punitive behavior by audio-visual displays. Science, 1962, 136 (Whole No. 3519), 872-873.

150. Wendt, Paul R. Factors Affecting Fixation Points in Viewing an Educational Motion Picture. Unpublished doctor's dissertation. University of Minnesota, 1948.

151. Williams, O. A study of the phenomenon of reminiscence. Journal of Experimental Psychology, 1926, 9, 368-387.

152. Yule, G. Udny, and Kendall, M. G. An Introduction to the Theory of Statistics, Fourteenth Edition. London: Charles Griffin, 1950.